

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/101759/>

Copyright and reuse:

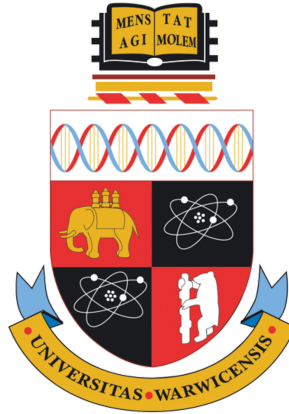
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



**Rates of Convergence for Statistical Limit Laws in
Deterministic Dynamical Systems**

by

Marios Antoniou

Thesis

Submitted to the University of Warwick
for the degree of

Doctor of Philosophy

Warwick Mathematics Institute

March 2018



Contents

ACKNOWLEDGEMENTS	iii
ABSTRACT	v
1 Introduction	1
2 Preliminaries	6
3 Martingale Approximations	11
3.1 The Koopman and Transfer Operators	11
3.2 An Example: The Doubling Map	13
3.3 General Framework for Maps	16
3.4 Martingale Decomposition	17
4 The Central Limit Theorem	20
4.1 Introduction and Statement of the Theorem	20
4.2 Central Limit Theorem via Lévy's Continuity Theorem	21
4.3 Martingale Theory	28
4.4 Central Limit Theorem via Martingales	32

5	Rate of Convergence in the Central Limit Theorem	39
5.1	Introduction	39
5.2	Rate of Convergence for Martingale Approximations	40
6	Weak Invariance Principle	45
6.1	Weak Convergence and Donsker's Invariance Principle	45
6.2	The Weak Invariance Principle for Martingale Approximations	47
7	Rate of convergence in the Weak Invariance Principle	55
7.1	Introduction and Statement of the Theorem	55
7.2	Error Rate in the WIP for martingale approximations	57
8	Fast-slow systems in Continuous Time	65
8.1	Introduction	65
8.2	Setting and Assumptions	69
8.3	Rates of Convergence in the Additive Noise case for Flows	70
8.4	Rates of Convergence in the Multiplicative Noise case for Flows	82
9	Fast-Slow systems in Discrete Time	85
9.1	Setting and Assumptions	85
9.2	Rates of Convergence in the Additive Noise case for Maps	86
9.3	Rates of Convergence in the Multiplicative Noise case for Maps	95
	BIBLIOGRAPHY	99

Acknowledgements

To my supervisor Professor Ian Melbourne I extend my wholehearted gratitude and appreciation. It may sound traditional and monotone to emphasise gratefulness towards one's supervisor. The simple fact is however that throughout my PhD studies he was always next to me, checking and correcting countless drafts of this thesis, patiently advising me and always thinking of the next step till we cross the "almost there" barrier. It was no surprise that he assumed no deadline on his mentoring role. He has rightfully earned my admiration and I am forever in his debt. I would like to sincerely thank Professor Richard Sharp for his many useful suggestions and comments on an earlier draft of the thesis, and, above all, for his always-friendly approach. Many thanks also to Professor Grigorios A. Pavliotis for useful suggestions and comments and for participating in this project. The Warwick Mathematics Institute offers a favourable and dynamic atmosphere for mathematical research. I am grateful for the opportunity granted to attend PhD studies at this fine institution. I acknowledge the generous financial support from European Advanced Grant (ERC AdG 320977). Lastly, but certainly not least, I would like to thank my family. My little sister Antria, one of the bravest people I know. My brother Constantinos for always checking up on me and for never being out of reach. My mother Litsa, for doing brilliantly what a mother does best and for patiently counting days all these years. Above all I want to thank my father for his never-ending support, for always being there when I needed him, for his countless sacrifices so that I can have the best future possible and for many, many other reasons.

Declaration

I declare that this thesis is my own work. I confirm that the thesis has not been submitted for a degree at another university.

ABSTRACT

We investigate rates of convergence in statistical limit theorems for observables of deterministic dynamical systems and the corresponding questions in homogenization of fast-slow systems. In particular we first use martingale approximations to review the Central Limit Theorem for ergodic stochastic processes under a general framework for expanding maps and retrieve the corresponding rate of convergence to a normal law. We then consider the functional central theorem under a general framework and obtain using a new method the corresponding rate of convergence to a Brownian motion. The main result of the thesis is establishing rates of convergence in homogenization for deterministic maps and multiplicative noise.

Dedicated to the memory of Χριστόφορος

Όμως υπάρχει τι το ανθρώπινον χωρίς ατέλεια;
Και τέλος πάντων, να, τραβούμ' εμπρός.

Κωνσταντίνος Καβάφης

Θαρσεῖν χρή... τάχ' αὔριον ἔσσετ' ἄμεινον...

Θεόκριτος

Chapter 1

Introduction

We deal with the classical problem in Ergodic Theory of understanding the statistical properties of typical orbits. It is well known that iterating a “sufficiently” chaotic map exhibits random behaviour leading to probabilistic limit laws. The starting point is Birkhoff’s ergodic theorem [10] (time averages starting from typical points converge to the space average) which is a generalisation of the strong law of large numbers.

Starting with work of [76, 68], results have been established in connection with the deviation from these averages, that is to say Central Limit Theorems for deterministic systems. One method for studying such limit theorems is the so-called “martingale approximation” method [31]. This method allows us to approach the problem in such a way that tools and results from martingale theory can be readily applied. We review this method and see how the rate of convergence for the Central Limit Theorem, i.e. the difference between the approximating expression and its limit, can be estimated using martingale techniques.

The Weak Invariance Principle [20] (also known as the functional Central Limit Theorem) is a far-reaching extension of the Central Limit Theorem whereby suitable scaled (in time and space) sums of random variables converge weakly to Brownian motion. This is also much-studied for deterministic dynamical

systems [29]. Under the conditions used in this thesis the Weak Invariance Principle was first proved by Hofbauer and Keller [38] using different techniques. A natural question is to ask about the convergence rate to Brownian motion in the Lévy-Prokhorov metric [67]. Here, very little seems to be known in the dynamical systems literature; we could only find one relevant paper [35] which is restricted to situations where there is a spectral gap (and well-illustrated for the case of Markov chains). Spectral methods are effective for proving Central Limit Theorems but less versatile than martingales ([34] and the references therein). Martingale techniques promise to be much more flexible and widely applicable, so one of the main contributions of this thesis is to present the first results on convergence rates in the Weak Invariance Principle using martingale limit theorems. Specifically, we apply a result of Kubilius [52].

Considerable attention is found, both in the mathematics and applications literature, in understanding how randomness can emerge from deterministic systems. A simple mechanism for emergent stochastic behaviour is via homogenization of multiscale systems, see for example [63]. In particular, there has been recently much interest in homogenization of fast-slow dynamical systems, leading in the limit to certain stochastic differential equations [18, 19, 55, 32, 47, 48]. Convergence is again in the sense of weak convergence so it makes sense to investigate convergence rates in the Lévy-Prokhorov metric. There are no previous results on such convergence rates. Hence the principal aim of this thesis concerns the investigation of error rates in homogenization for deterministic systems using martingale techniques. In particular, we build on the work of [55, 32] which establishes homogenization results under the assumption that the slow dynamics is one-dimensional (in the additive and multiplicative noise cases) for very general fast dynamics.

The structure of the thesis is as follows:

In chapters 2 and 3 we review some basic definitions and results necessary to establish notation and background and to prepare the way for the rest of the thesis. In chapter 3 we describe a general framework under the setting of which the rates of convergence in the statistical limit laws in this thesis are deduced.

By way of illustration we consider the case where the relevant dynamics is driven by the doubling map. This general framework, following Field et al [29], is used to derive martingale decompositions for Hölder continuous observables satisfying properties that hold true for uniformly expanding maps. The martingale approximation method following a standard argument in Gordin [31], as shown in [29], expresses Hölder observables as the sum of a martingale part and an asymptotically negligible coboundary. Then various probabilistic limit theorems can be called upon to show that partial sums of these observables satisfy Central Limit Theorems and invariance principles. This feeds into chapters 4 and 6.

The Central Limit Theorem in chapter 4 is well known. We follow two standard proofs, reviewing arguments and methodology which are referred to for establishing theorems in later chapters. The first proof uses multiplicative sequences following McLeish [56]. The second proof of the Central Limit Theorem exploits the martingale part in the decomposition described above and utilises standard martingale theory concepts. In particular, the martingale Central Limit Theorem, established independently by Billingsley [7] and Ibragimov [40] is readily applicable.

In chapter 5 we obtain the rate of convergence in the Central Limit Theorem using the approximation of the partial sums of stationary stochastic processes under our general framework. Such results are known using different techniques [33]. Several results on the error rate in martingale Central Limit Theorems have been obtained under a variety of assumptions [11, 36, 37, 43, 58]. We conclude the uniform bound on the speed of convergence using results for martingale increments obtained in [58]. As we see here, and in later chapters, the telescoping coboundary does not affect the error rate.

Chapter 6 deals with the functional Central Limit Theorem under the general framework. Invariance principles for dependent random variables, extending Donsker's invariance principle [20] for independent and identically distributed random variables, are well known [6, 9, 36, 57]. The Weak Invariance Principle for martingale approximations follows in a straightforward manner.

In chapter 7 we deduce the rate of convergence in the Weak Invariance Principle in the context of the general framework using a new method. Starting from a result by Kubilius [52] on the error rate in the Weak Invariance Principle for martingale difference arrays we employ martingale theory arguments to conclude the result.

In chapters 8 and 9 we obtain error rates in homogenization of fast-slow systems first for continuous time systems (differential equations) in Chapter 8, and then for discrete time systems (maps) in Chapter 9.

Remark. Most of this thesis, with the exception of chapter 8, is written in the context of discrete time dynamical systems generated by uniformly expanding maps. In ergodic theory, it is well known (see for example [60]) that proving statistical limit laws is easier for discrete time dynamical systems than for continuous time systems. Hence in chapters 3 to 7 (Central Limit Theorems and invariance principles) we focus entirely on the discrete time setting. However, for fast-slow systems the situation is reversed: given good understanding of the fast dynamics, the homogenization problem is easier for continuous time. Hence we first prove convergence rates for homogenization for continuous time in chapter 8; here we have to assume that the fast dynamics satisfies strong statistical properties. Then in chapter 9 we cover the discrete time case under the assumption that the fast dynamics is uniformly expanding – this time the required assumptions on the fast dynamics follow from the earlier chapters in this thesis.

While this thesis was in the process of being written, new advances on martingale approximation methods for dynamical systems were obtained in [51]. When combined with [51], the methods in this thesis apply to large classes of nonuniformly expanding and nonuniformly hyperbolic dynamical systems, moving considerably beyond the uniformly expanding dynamical systems considered here. Given time constraints, this thesis focuses on developing new probabilistic results as far as possible in the simplest dynamical setting (uniformly expanding maps) rather than attempting to incorporate recent developments on the dynamical systems theory [51]. A joint paper, Antoniou &

Melbourne [1], is in preparation and combines the methods in the thesis with those in [51]. This paper is not part of the PhD thesis.

The main new results in this thesis are the ones on convergence rates in the Weak Invariance Principle (chapter 7) and convergence rates for homogenization of fast-slow dynamical systems (chapters 8 and 9). The earlier chapters serve as an exposition of known results and methods as background and preparation for the main results (though the material in chapter 5 on using martingale approximation to prove convergence rates in the Central Limit Theorem seems not to be readily available in the literature).

Chapter 2

Preliminaries

In this chapter we gather together some mathematical preliminaries and establish notation. Some further background and tools will be presented as needed throughout the thesis.

We begin by considering a *probability space*: a mathematical triplet $(\Lambda, \mathcal{F}, \mu)$ such that

- The set Λ is the *sample space*.
- \mathcal{F} is a σ -*algebra* on the space Λ , i.e. a collection of subsets of Λ with $\emptyset \in \mathcal{F}$ and which is closed under the set operations of complement and union of countably many sets.
- μ is a *probability measure* i.e. a map $\mu : \mathcal{F} \rightarrow [0, 1]$ with $\mu(\Lambda) = 1$ which is countably additive: If $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$ is any sequence of pairwise disjoint sets then

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A pair (Λ, \mathcal{F}) where Λ is a set and \mathcal{F} a σ -algebra on Λ is called a *measurable space*. The *Borel* σ -algebra $\mathcal{B}(\Lambda)$ on a (topological) space Λ is the σ -*algebra*

generated by the collection of all open sets,

$$\mathcal{B}(\Lambda) = \sigma(\{A \subseteq \Lambda \mid A \text{ is open}\}) .$$

Let (Λ, \mathcal{F}) and (S, \mathcal{S}) be measurable spaces. A function $f : \Lambda \rightarrow S$ is called $(\mathcal{F}, \mathcal{S})$ -measurable (or just measurable if \mathcal{F} and \mathcal{S} are understood) if $f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{S}$. Such a measurable function is called an (S, \mathcal{S}) -valued *random variable* (or just random variable). Two real random variables X and Y are equal in distribution (or equal in law), written $X =_d Y$, if they have the same distribution functions, $\mu(X \leq x) = \mu(Y \leq x)$ for all x . In general, $X =_d Y$ if $\mu(X \in A) = \mu(Y \in A)$ for all Borel sets A .

We say that f is *integrable* if we have $\int |f| d\mu < \infty$. The set of all integrable functions for μ is called $\mathcal{L}^1(\Lambda, \mathcal{F}, \mu)$. More generally, for $1 \leq p < \infty$ we denote by $\mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$ the space of all measurable functions $f : \Lambda \rightarrow \mathbb{R}$ such that $\|f\|_p < \infty$ where

$$\|f\|_p = \left(\int_{\Lambda} |f|^p d\mu \right)^{1/p} .$$

A measurable function f is called *essentially bounded* if and only if for some $M < \infty$, $|f| \leq M$ almost everywhere. We denote the space of essentially bounded real functions by $\mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$. For any f we write

$$\|f\|_\infty = \inf\{M \mid |f| \leq M \text{ a.e.}\} .$$

If f is continuous on $[0, 1]$ and μ is Lebesgue measure then $\|f\|_\infty = \sup|f|$.

Let $T : \Lambda \rightarrow \Lambda$ be a \mathcal{F} -measurable transformation i.e. it satisfies $T^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{F}$ where

$$T^{-1}(A) = \{x \in \Lambda : T(x) \in A\} .$$

Definition 2.1. We say that T is a *measure-preserving transformation* or, equivalently, that μ is a *T -invariant measure*, if we have $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{F}$.

We have the following characterization of invariant measures in terms of integrable functions e.g. [73].

Lemma 2.2. *A map $T : (\Lambda, \mathcal{F}, \mu) \rightarrow (\Lambda, \mathcal{F}, \mu)$ is measure-preserving if and only if for all $v \in \mathcal{L}^1(\Lambda, \mathcal{F}, \mu)$ we have*

$$\int v \circ T d\mu = \int v d\mu. \quad (2.1)$$

Let v be measurable and $A \in \mathcal{B}(\mathbb{R})$. Then $g = \mathbb{1}_{v \in A}$ is measurable and it is also integrable. By (2.1) we have $\int g \circ T d\mu = \int g d\mu$. So

$$\begin{aligned} \mu(v \circ T \in A) &= \int \mathbb{1}_{v \circ T \in A} d\mu = \int \mathbb{1}_{v \in A} \circ T d\mu = \int g \circ T d\mu \\ &= \int g d\mu = \int \mathbb{1}_{v \in A} = \mu(v \in A). \end{aligned}$$

Hence for a measure-preserving map T and an integrable function $v : \Lambda \rightarrow \mathbb{R}$ we have that v and $v \circ T$ are identically distributed,

$$\mu(y \in \Lambda : v(y) < b) = \mu(y \in \Lambda : v(Ty) < b) \quad \text{for all } b \in \mathbb{R}.$$

Since by (2.1) we have inductively that

$$\int v \circ T^k d\mu = \int v \circ T^{k-1} \circ T d\mu = \dots = \int v d\mu \quad \text{for all } k \in \mathbb{N},$$

it follows that $v, v \circ T, v \circ T^2, \dots$ form a sequence of identically distributed random variables. They are not usually independent, but as we shall see in the next chapter for a concrete example, their correlation decays asymptotically fast. It is natural then to ask about the limiting behaviour of various averages over time. Instead of requiring independence of the random variables we consider a weaker condition of which independence is a special case. This condition is stationarity [14].

Definition 2.3. A stochastic process $\{X_i\}_{i \in \mathbb{N}}$ on $(\Lambda, \mathcal{F}, \mu)$ is stationary if for all $j \geq 1$, $n \in \mathbb{N}$,

$$\mathbb{P}[X_{n+1} \in B_1, X_{n+2} \in B_2, \dots, X_{n+j} \in B_j] = \mathbb{P}[X_1 \in B_1, X_2 \in B_2, \dots, X_j \in B_j]$$

for every $B_1, B_2, \dots, B_j \in \mathcal{F}$.

It is standard, [14], that if T is measure-preserving on $(\Lambda, \mathcal{F}, \mu)$ and v is a random variable on (Λ, \mathcal{F}) then the sequence $\{v \circ T^n\}_{n \geq 0}$ is a stationary sequence of random variables.

Definition 2.4. We say that μ , (respectively T) is ergodic with respect to a measure-preserving transformation T , (respectively T -invariant measure μ) if for $A \in \mathcal{F}$ we have $T^{-1}(A) = A$ implies $\mu(A) \in \{0, 1\}$, i.e. there are no non-trivial invariant sets.

If T is an ergodic map we then have a strong law of large numbers for stationary stochastic processes. This is the celebrated Birkhoff's Ergodic Theorem [10, 64].

Theorem 2.5. (Birkhoff's Ergodic Theorem) Let $(\Lambda, \mathcal{F}, \mu)$ be a probability space and assume that the measure-preserving transformation $T : \Lambda \rightarrow \Lambda$ is ergodic. If $v \in \mathcal{L}^1(\Lambda, \mathcal{F}, \mu)$ then

$$\frac{1}{n} \sum_{i=0}^{n-1} v \circ T^i \rightarrow \int v d\mu \quad a.e.$$

$$i.e. \mu(y \in \Lambda : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} v \circ T^i(y) \neq \int v d\mu) = 0.$$

Now assume that $\int v d\mu = 0$ and denote by v_n the n 'th partial sum i.e.

$$v_n = \sum_{i=0}^{n-1} v \circ T^i.$$

The questions that we address in this thesis deal with convergence behaviour for observables of deterministic dynamical systems and for homogenization of

fast-slow systems. In particular our aim is to obtain the corresponding rates of convergence. For Birkhoff sums v_n we ask: What is $\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} v_n$? How fast does it converge to that limit? For more complicated functionals of these Birkhoff sums we investigate convergence rates to a Brownian motion [4, 41, 45, 62, 75].

Definition 2.6. *We say that a stochastic process $W = \{W_t\}_{t \geq 0}$ is a Brownian motion with variance $\sigma^2 > 0$ if*

- $W_0 = 0$ a.s.
- for all $t_0 \leq t_1 \leq \dots \leq t_n$ we have that $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ are independent. (independent increments)
- $W_t - W_s =_d W_{t-s}$ for $0 \leq s < t$. (stationary increments)
- $W_t \sim \mathcal{N}(0, \sigma^2 t)$ for $t \geq 0$. (normality)
- $t \mapsto W_t$ is continuous a.s. (continuous sample paths)

We will refer to a stochastic process W with $\sigma^2 = 1$ in Definition 2.6 as a standard Brownian motion.

Chapter 3

Martingale Approximations

3.1 The Koopman and Transfer Operators

We proceed by considering some key tools in our study of statistical limit laws. Firstly, let us introduce the *composition* or *Koopman operator* [50] and, its adjoint, the *transfer* or *Perron-Frobenius operator* [3]. We assume throughout this thesis that $(\Lambda, \mathcal{F}, \mu)$ is a probability space.

Definition 3.1. Let $T : \Lambda \rightarrow \Lambda$ be a measure-preserving transformation. The Koopman operator $\mathcal{U} : \mathcal{L}^1(\Lambda, \mathcal{F}, \mu) \rightarrow \mathcal{L}^1(\Lambda, \mathcal{F}, \mu)$ associated with T is defined by

$$\mathcal{U}v = v \circ T.$$

Definition 3.2. Let $T : \Lambda \rightarrow \Lambda$ be a measure-preserving transformation. The unique operator $\mathcal{P} : \mathcal{L}^p(\Lambda, \mathcal{F}, \mu) \rightarrow \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$, $p \geq 1$, such that for $v \in \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$ we have

$$\int_{\Lambda} (\mathcal{P}v)w \, d\mu = \int_{\Lambda} v(\mathcal{U}w) \, d\mu \quad \text{for all } w \in \mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$$

is called the transfer operator associated with T .

That is to say, the transfer operator is dual to the Koopman operator in the sense that $\langle \mathcal{P}v, w \rangle = \langle v, \mathcal{U}w \rangle$ where $\langle f, g \rangle = \int fg d\mu$. Some standard properties satisfied by these two operators are given below (e.g. [13]) .

Proposition 3.3. *Let $T : \Lambda \rightarrow \Lambda$ be a measure-preserving transformation on the probability space $(\Lambda, \mathcal{F}, \mu)$ and \mathcal{U} and \mathcal{P} be the associated Koopman and transfer operators respectively as defined above. The following hold true:*

(a) $\mathcal{U}1 = 1, \mathcal{P}1 = 1$.

(b) $\|\mathcal{U}v\|_p = \|v\|_p$ and $\|\mathcal{P}v\|_p \leq \|v\|_p$ for all $v \in \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$.

(c) $\int_{\Lambda} \mathcal{U}v \mathcal{U}w d\mu = \int_{\Lambda} vw d\mu$ for all $v \in \mathcal{L}^1(\Lambda, \mathcal{F}, \mu)$, $w \in \mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$.

(d) $\mathcal{P}\mathcal{U}v = v$ for all $v \in \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$.

(e) $\int_{\Lambda} (\mathcal{P}^n v)w d\mu = \int_{\Lambda} v(\mathcal{U}^n w) d\mu$ for all $n \geq 1$, for all $v \in \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$, $w \in \mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$.

(f) $\int_{\Lambda} \mathcal{P}^n v d\mu = \int_{\Lambda} v d\mu$ for all $n \geq 1$, for all $v \in \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$.

(g) $\int_{\Lambda} v d\mu = 0 \implies \int_{\Lambda} \mathcal{P}v d\mu = 0$ for all $v \in \mathcal{L}^p(\Lambda, \mathcal{F}, \mu)$.

Proof: These properties follow from the definitions of \mathcal{P} and \mathcal{U} as well as T -invariance for μ , e.g. for (d) observe that for an arbitrary w we have

$$\int_{\Lambda} \mathcal{P}(\mathcal{U}v)(w) d\mu = \int_{\Lambda} \mathcal{U}v \mathcal{U}w d\mu = \int_{\Lambda} (v \circ T)(w \circ T) d\mu = \int_{\Lambda} (vw) d\mu.$$

For (e) note that

$$\int_{\Lambda} (\mathcal{P}^n v)w d\mu = \int_{\Lambda} (\mathcal{P}(\mathcal{P}^{n-1}v))w d\mu = \int_{\Lambda} (\mathcal{P}^{n-1}v)(\mathcal{U}w) d\mu = \cdots = \int_{\Lambda} v(\mathcal{U}^n w) d\mu.$$

■

3.2 An Example: The Doubling Map

We now look at a concrete example. Let $\Lambda = [0, 1]$ and consider the *doubling map*, figure 3.1, defined by

$$T : \Lambda \rightarrow \Lambda, \quad T(x) = 2x \bmod 1.$$

This is a Lebesgue measure-preserving and ergodic transformation [13].

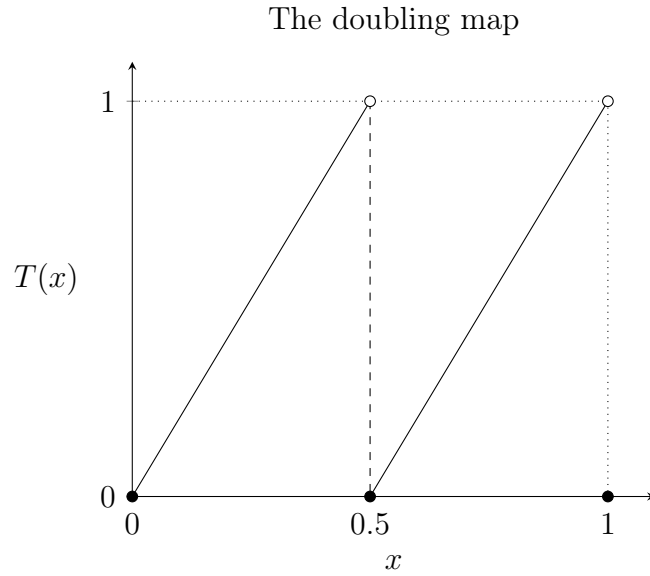


Figure 3.1: The doubling map

Proposition 3.4. *The transfer operator for the doubling map is given by*

$$(\mathcal{P}v)(x) = \frac{1}{2}v\left(\frac{x}{2}\right) + \frac{1}{2}v\left(\frac{x+1}{2}\right).$$

Proof: Observe that

$$\begin{aligned} \int (\mathcal{P}v)w \, d\mu &= \int_0^1 v(x)w(2x \bmod 1) \, dx \\ &= \int_0^{1/2} v(x)w(2x) \, dx + \int_{1/2}^1 v(x)w(2x-1) \, dx \\ &= \frac{1}{2} \int_0^1 v\left(\frac{y}{2}\right) w(y) \, dy + \frac{1}{2} \int_0^1 v\left(\frac{y+1}{2}\right) w(y) \, dy. \end{aligned}$$

■

Lipschitz Space: Let (Λ, ρ) be a metric space. Recall that for a bounded real function f on Λ the Lipschitz seminorm is defined by

$$\text{Lip}(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x, y)}.$$

Let $\|\cdot\|_\infty$ denote the supremum norm, $\|f\|_\infty = \sup_x |f(x)|$. The family of all bounded real-valued Lipschitz continuous functions is denoted by

$$\text{Lip}(\Lambda, \rho) = \left\{ f : \Lambda \rightarrow \mathbb{R} \mid \|f\|_{\text{Lip}} < \infty \right\}$$

where $\|\cdot\|_{\text{Lip}} = \text{Lip}(\cdot) + \|\cdot\|_\infty$. It is standard that $(\text{Lip}(\Lambda, \rho), \|\cdot\|_{\text{Lip}})$ is a Banach space.

Proposition 3.5. *Consider the doubling map and let v be a Lipschitz continuous observable. We have*

$$\text{Lip}(\mathcal{P}^n v) \leq \left(\frac{1}{2}\right)^n \text{Lip}(v).$$

Proof: Observe that

$$\begin{aligned} |(\mathcal{P}v)(x) - (\mathcal{P}v)(y)| &\leq \frac{1}{2} \left| v\left(\frac{x}{2}\right) - v\left(\frac{y}{2}\right) \right| + \frac{1}{2} \left| v\left(\frac{x+1}{2}\right) - v\left(\frac{y+1}{2}\right) \right| \\ &\leq \frac{1}{2} \text{Lip}(v) \left| \frac{x}{2} - \frac{y}{2} \right| + \frac{1}{2} \text{Lip}(v) \left| \frac{x}{2} - \frac{y}{2} \right| = \frac{1}{2} \text{Lip}(v) |x - y|. \end{aligned}$$

This proves the result for $n = 1$ and the general case follows by induction. ■

Theorem 3.6. *Consider the doubling map and an observable $v : \Lambda \rightarrow \mathbb{R}$ with $\int v d\mu = 0$ and Lipschitz continuous. Then it holds true that*

$$\|\mathcal{P}^n v\|_{\text{Lip}} \leq \frac{1}{2^{n-1}} \text{Lip}(v).$$

Proof: If g is a real-valued bounded function on Λ with $\text{Lip}(g) < \infty$ then it is clear that

$$\left\| g - \int g d\mu \right\|_{\infty} \leq \sup_{y,z} |g(y) - g(z)| \leq \text{Lip}(g) \text{diam}(\Lambda).$$

Using Proposition 3.3 (f) and since $\text{diam}(\Lambda) = 1$ and $\int v d\mu = 0$ we have

$$\|\mathcal{P}^n v\|_{\infty} = \left\| \mathcal{P}^n v - \int v d\mu \right\|_{\infty} = \left\| \mathcal{P}^n v - \int \mathcal{P}^n v d\mu \right\|_{\infty} \leq \text{Lip}(\mathcal{P}^n v).$$

Hence we conclude by Proposition 3.5 and the definition of $\|\cdot\|_{\text{Lip}}$ that

$$\|\mathcal{P}^n v\|_{\text{Lip}} = \|\mathcal{P}^n v\|_{\infty} + \text{Lip}(\mathcal{P}^n v) \leq 2 \text{Lip}(\mathcal{P}^n v) \leq 2 \frac{1}{2^n} \text{Lip}(v).$$

■

Proposition 3.7. (Decay of Correlations) *Let T be the doubling map on $\Lambda = [0, 1]$. Assume that $v : \Lambda \rightarrow \mathbb{R}$ is a Lipschitz continuous observable and w is integrable. Let $C(v, w)$ be the correlation function for v and w ,*

$$C(v, w) = \int_{\Lambda} v w \circ T^n d\mu - \int_{\Lambda} v d\mu \int_{\Lambda} w d\mu$$

Then $|C(v, w)| \leq \frac{1}{2^n} \text{Lip}(v) \|w\|_1$.

Proof: We have

$$C(v, w) = \int_{\Lambda} \mathcal{P}^n v w d\mu - \int_{\Lambda} v d\mu \int_{\Lambda} w d\mu = \int_{\Lambda} \left(\mathcal{P}^n v - \int_{\Lambda} v \right) w d\mu.$$

Notice that by Proposition 3.3 part (f) we have $\int_{\Lambda} \mathcal{P}^n v d\mu = \int_{\Lambda} v d\mu$. Since v is Lipschitz we have $\|\mathcal{P}^n v - \int_{\Lambda} \mathcal{P}^n v d\mu\|_{\infty} \leq \text{Lip}(\mathcal{P}^n v) \text{diam}(\Lambda) = \text{Lip}(\mathcal{P}^n v)$. By Proposition 3.5 we have $\text{Lip}(\mathcal{P}^n v) \leq \left(\frac{1}{2}\right)^n \text{Lip}(v)$. Hence we conclude that

$$\begin{aligned} |C(v, w)| &\leq \left\| \mathcal{P}^n v - \int_{\Lambda} v d\mu \right\|_{\infty} \|w\|_1 = \left\| \mathcal{P}^n v - \int_{\Lambda} \mathcal{P}^n v d\mu \right\|_{\infty} \|w\|_1 \\ &\leq \text{Lip}(\mathcal{P}^n v) \|w\|_1 \leq \frac{1}{2^n} \text{Lip}(v) \|w\|_1 . \end{aligned}$$

■

3.3 General Framework for Maps

Recall the definition of an expanding map [46]:

Definition 3.8. *A continuous map $f : X \rightarrow X$ where (X, d) is a metric space, is called expanding if for some $k > 1$, $\epsilon > 0$ and every $x, y \in X$ with $x \neq y$ and $d(x, y) < \epsilon$ we have*

$$d(f(x), f(y)) > kd(x, y) .$$

The result for the doubling map in Theorem 3.6 is a special case of a general property that holds true for a large class of expanding maps. In particular we assume that the underlying dynamical setting in the discrete time cases considered in this thesis satisfies the following:

General Framework (H): Let Λ be a compact metric space with Borel σ -algebra $\mathcal{B}(\Lambda)$ and probability measure μ . Let $T : \Lambda \rightarrow \Lambda$ be a measure-preserving, ergodic transformation. We assume that there exist constants $C > 0$, $\gamma \in (0, 1)$ such that

$$\|\mathcal{P}^n v\|_{\alpha} \leq C \gamma^n \|v\|_{\alpha}$$

for all $v \in \mathcal{C}^\alpha(\Lambda)$ with $\int_\Lambda v d\mu = 0$ and all $n \geq 1$. As standard, $\mathcal{C}^\alpha(\Lambda)$ is the Banach space of Hölder continuous functions with exponent α , equipped with the norm given by $\|\cdot\|_\alpha = \|\cdot\|_\infty + |\cdot|_\alpha$. Here

$$|v|_\alpha = \sup_{x \neq y} \frac{|v(x) - v(y)|}{\rho(x, y)^\alpha}$$

is the Hölder seminorm.

3.4 Martingale Decomposition

Martingale approximation methods were introduced by Gordin [31] in 1969 and have been utilised for investigating various Central Limit Theorems ever since. We follow [29] in using martingale approximations to study statistical laws for dynamical systems under the general framework of section 3.3.

Theorem 3.9. (Martingale Approximation) *Assume that T satisfies the General Framework (H) and $v \in \mathcal{C}^\alpha(\Lambda)$ with $\int_\Lambda v d\mu = 0$. Then there exists $\chi \in \mathcal{C}^\alpha(\Lambda)$, $m \in \mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$ such that v has the decomposition*

$$v = m + \chi \circ T - \chi \tag{3.1}$$

with $m \in \text{Ker}(\mathcal{P})$ and $\int_\Lambda m d\mu = 0$.

Proof: Set $\chi = \sum_{n=1}^\infty \mathcal{P}^n v$. By the General Framework (H) we have that χ converges absolutely in the Banach space of Hölder continuous functions and is therefore Hölder. In particular, $\chi \in \mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$. Define $m = v - \chi \circ T + \chi$. Then $\|m\|_\infty \leq \|v\|_\infty + 2\|\chi\|_\infty$ so $m \in \mathcal{L}^\infty(\Lambda, \mathcal{F}, \mu)$. Moreover,

$$0 = \int_\Lambda v d\mu = \int_\Lambda m d\mu + \int_\Lambda \chi \circ T d\mu - \int_\Lambda \chi d\mu = \int_\Lambda m d\mu$$

where the last equality follows from T -invariance for μ . Lastly, observe that

$$\chi - \mathcal{P}\chi = \sum_{n=1}^{\infty} \mathcal{P}^n v - \sum_{n=1}^{\infty} \mathcal{P}^{n+1} v = \sum_{n=1}^{\infty} \mathcal{P}^n v - \sum_{n=2}^{\infty} \mathcal{P}^n v = \mathcal{P}v. \quad (3.2)$$

Thus applying the transfer operator to (3.1) we conclude that

$$\mathcal{P}m = \mathcal{P}v - \mathcal{P}\mathcal{U}\chi + \mathcal{P}\chi = \mathcal{P}v - \chi + \mathcal{P}\chi \stackrel{(3.2)}{=} \mathcal{P}v - \mathcal{P}v = 0$$

where the second equality follows from Proposition 3.3 (d). ■

Remark 3.10. *The monograph [60] presents in more detail the general framework for martingale approximations (see in particular Proposition 1.2 therein). More on necessary and sufficient conditions for v to have the decomposition (3.1) can be found in [85] and the references therein. The idea is that m behaves in some sense, to be made precise in the next chapter, like a martingale and this is perturbed by the asymptotically negligible coboundary $\chi \circ T - \chi$. Hence, approximating a stationary ergodic sequence sufficiently closely by a martingale reduces the limit theorem problem to a simplified one for martingales which is easier to deal with. Note that if v is not mean zero we can replace it by an observable $\hat{v} = v - \int v d\mu$ and deduce the theory for \hat{v} .*

Now even though our random variables $m, m \circ T, m \circ T^2, \dots$ are not independent we have the following property at hand:

Proposition 3.11. *Let m, T be as in Theorem 3.9. The family of random variables $\{m \circ T^i\}_{i \geq 0}$ is multiplicative (or orthogonal):*

$$\mathbb{E}_\mu \left(\prod_{i=1}^k m \circ T^{j_i} \right) = 0 \quad \forall 0 \leq j_1 < j_2 < \dots < j_k, k \geq 1. \quad (3.3)$$

Proof: We have from Theorem 3.9 that μ is T -invariant, $m \in \text{Ker}(\mathcal{P})$ and $\int_\Lambda m d\mu = 0$. First note that for $k = 1$ it is trivial that $\int_\Lambda m \circ T^{j_1} d\mu =$

$\int_{\Lambda} m \, d\mu = 0$. For $k \geq 2$ observe that

$$\begin{aligned}
 & \int_{\Lambda} m \circ T^{j_k} m \circ T^{j_{k-1}} \dots m \circ T^{j_1} \, d\mu \\
 &= \int_{\Lambda} \left(m \left(m \circ T^{j_k-j_1} m \circ T^{j_{k-1}-j_1} \dots m \circ T^{j_2-j_1} \right) \right) \circ T^{j_1} \, d\mu \\
 &= \int_{\Lambda} m \left(m \circ T^{j_k-j_1} m \circ T^{j_{k-1}-j_1} \dots m \circ T^{j_2-j_1} \right) \, d\mu \\
 &= \int_{\Lambda} m \left[\left(m \circ T^{j_k-j_1-1} m \circ T^{j_{k-1}-j_1-1} \dots m \circ T^{j_2-j_1-1} \right) \circ T \right] \, d\mu \\
 &= \int_{\Lambda} \mathcal{P}m \left(m \circ T^{j_k-j_1-1} m \circ T^{j_{k-1}-j_1-1} \dots m \circ T^{j_2-j_1-1} \right) \, d\mu = 0.
 \end{aligned}$$

■

Chapter 4

The Central Limit Theorem

4.1 Introduction and Statement of the Theorem

Let us begin by recalling the classical Central Limit Theorem (Lindeberg–Lévy theorem) [16, 25]: If $\{X_i\}_{i \in \mathbb{N}}$ is a sequence of independent and identically distributed random variables (i.i.d. r.v.'s) with $\mathbb{E}[X_i] = 0$ and $\text{Var}(X_i) = \mathbb{E}[X_i^2] = \sigma^2 < \infty$ for all $i \geq 1$ then it holds true that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d \mathcal{N}(0, \sigma^2) \quad \text{as } n \rightarrow \infty$$

where $\mathcal{N}(0, \sigma^2)$ is the normal distribution for a zero-mean random variable with variance equal to σ^2 . This means that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \leq t \right] = \mathbb{P}[Y \leq t]$$

where $Y \sim \mathcal{N}(0, \sigma^2)$ i.e. its distribution function is given by

$$\mathbb{P}[Y \leq t] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2\sigma^2}} ds.$$

In this chapter we consider the sequence of identically distributed but not independent random variables $\{v \circ T^i\}_{i \geq 0}$ where the underlying dynamics throughout the chapter satisfies the General Framework **(H)**. Let $v_n = \sum_{i=0}^{n-1} v \circ T^i$. We show that the distribution of $n^{-1/2}v_n$ is asymptotically normal.

In particular we prove the following:

Theorem 4.1. (Central Limit Theorem) *Assume that T satisfies the General Framework **(H)** and let v be Hölder continuous and mean zero. Let $Y \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = \int_{\Lambda} m^2 d\mu$. The Central Limit Theorem holds:*

$$\lim_{n \rightarrow \infty} \mu \left(y \in \Lambda : \frac{1}{\sqrt{n}} v_n(y) \leq t \right) = \mathbb{P}[Y \leq t] .$$

We give two proofs of Theorem 4.1. The proof in section 4.2 below follows closely McLeish [56]. For the second proof in section 4.4 we follow the formulation presented in [29] Remark 3.12 utilising martingale theory concepts presented in section 4.3.

4.2 Central Limit Theorem via Lévy's Continuity Theorem

We begin with the following elementary calculation.

Lemma 4.2. *Consider the martingale approximation (3.1) in Theorem 3.9. We have $v_n = m_n + \chi \circ T^n - \chi$.*

Proof: Observe that

$$\begin{aligned}
 \sum_{i=0}^{n-1} (\chi \circ T - \chi) \circ T^i &= \sum_{i=0}^{n-1} \chi \circ T^{i+1} - \sum_{i=0}^{n-1} \chi \circ T^i \\
 &= \left(\chi \circ T^n + \sum_{i=1}^{n-1} \chi \circ T^i \right) - \left(\sum_{i=1}^{n-1} \chi \circ T^i + \chi \right) \\
 &= \chi \circ T^n - \chi.
 \end{aligned} \tag{4.1}$$

Hence summing (3.1) we have by (4.1) that

$$\begin{aligned}
 v_n &= \sum_{i=0}^{n-1} m \circ T^i + \sum_{i=0}^{n-1} (\chi \circ T - \chi) \circ T^i \\
 &= m_n + \chi \circ T^n - \chi.
 \end{aligned}$$

■

Next we show that it suffices to prove the CLT for $\{m \circ T^i\}_{i \geq 0}$. We only use that $\chi \in \mathcal{L}^2(\Lambda, \mathcal{F}, \mu)$.

Lemma 4.3. *Let $\chi \in \mathcal{L}^2(\Lambda, \mathcal{F}, \mu)$ and Y be a random variable. Then*

$$\frac{1}{\sqrt{n}} v_n \rightarrow_d Y \text{ if and only if } \frac{1}{\sqrt{n}} m_n \rightarrow_d Y.$$

Proof: Since $\chi \in \mathcal{L}^2(\Lambda, \mathcal{F}, \mu)$ we have that $\chi^2 \in \mathcal{L}^1(\Lambda, \mathcal{F}, \mu)$ and by Birkhoff's Ergodic Theorem

$$\begin{aligned}
 \left(\frac{1}{\sqrt{n}} \chi \circ T^n \right)^2 &= \frac{1}{n} \chi^2 \circ T^n = \frac{n+1}{n} \left(\frac{1}{n+1} \sum_{i=0}^n \chi^2 \circ T^i \right) - \frac{1}{n} \sum_{i=0}^{n-1} \chi^2 \circ T^i \\
 &\rightarrow \int_{\Lambda} \chi^2 d\mu - \int_{\Lambda} \chi^2 d\mu = 0 \quad \text{a.e.}
 \end{aligned}$$

i.e. $\frac{1}{\sqrt{n}}\chi \circ T^n \rightarrow 0$ almost everywhere. Since by Lemma 4.2 we have

$$\frac{1}{\sqrt{n}}v_n - \frac{1}{\sqrt{n}}m_n = \frac{1}{\sqrt{n}}(\chi \circ T^n - \chi)$$

the result follows. ■

Hence the effect of $\chi \circ T^n - \chi$ disappears under suitable rescaling and we study the CLT problem for $\{m \circ T^i\}_{i \geq 0}$, where m is as in Theorem 3.9 and Proposition 3.11. We proceed by identifying the variance parameter σ^2 .

Proposition 4.4. *The following hold true:*

1. *The limit $\sigma^2 = \lim_{n \rightarrow \infty} \int_{\Lambda} \left(\frac{1}{\sqrt{n}}v_n\right)^2 d\mu$ exists.*
2. *$\lim_{n \rightarrow \infty} \int_{\Lambda} \left(\frac{1}{\sqrt{n}}m_n\right)^2 d\mu$ exists.*

Moreover, the two limits are equal to $\sigma^2 = \int_{\Lambda} m^2 d\mu$.

Proof: For 2. and the final statement, observe that

$$\begin{aligned} \int_{\Lambda} \left(\frac{1}{\sqrt{n}}m_n\right)^2 d\mu &= \frac{1}{n} \int_{\Lambda} \left(\sum_{i=0}^{n-1} m \circ T^i\right)^2 d\mu = \frac{1}{n} \sum_{i,j=0}^{n-1} \int_{\Lambda} m \circ T^i m \circ T^j d\mu \\ &\stackrel{(3.3)}{=} \frac{1}{n} \sum_{i=0}^{n-1} \int_{\Lambda} m \circ T^i m \circ T^i d\mu = \frac{1}{n} \sum_{i=0}^{n-1} \int_{\Lambda} (m \circ T^i)^2 d\mu = \frac{1}{n} \sum_{i=0}^{n-1} \int_{\Lambda} m^2 \circ T^i d\mu \\ &\stackrel{(2.1)}{=} \int_{\Lambda} m^2 d\mu = \sigma^2. \end{aligned}$$

Using Minkowski's inequality we have

$$\begin{aligned} \left| \|n^{-1/2}v_n\|_2 - \|n^{-1/2}m_n\|_2 \right| &\leq \|n^{-1/2}(v_n - m_n)\|_2 \\ &= \|n^{-1/2}(\chi \circ T^n - \chi)\|_2 \leq 2n^{-1/2} \|\chi\|_2. \end{aligned}$$

Since $\chi \in \mathcal{L}^2(\Lambda, \mathcal{F}, \mu)$ we conclude that

$$\lim_{n \rightarrow \infty} \int_{\Lambda} \left(\frac{1}{\sqrt{n}} v_n \right)^2 d\mu = \lim_{n \rightarrow \infty} \int_{\Lambda} \left(\frac{1}{\sqrt{n}} m_n \right)^2 d\mu = \sigma^2.$$

■

For completeness we include the next two results concerning the variance σ^2 .

Proposition 4.5. $\sigma^2 = \int_{\Lambda} v^2 d\mu + 2 \sum_{n=1}^{\infty} \int_{\Lambda} v v \circ T^n d\mu$.

Proof:

$$\begin{aligned} \int_{\Lambda} v_n^2 d\mu &= \int_{\Lambda} \sum_{j=0}^{n-1} v \circ T^j \sum_{k=0}^{n-1} v \circ T^k d\mu = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \int_{\Lambda} v \circ T^j v \circ T^k d\mu \\ &= \sum_{j=0}^{n-1} \int_{\Lambda} v \circ T^j v \circ T^j d\mu + 2 \sum_{0 \leq j < k \leq n-1} \int_{\Lambda} v \circ T^j v \circ T^k d\mu \\ &= n \int_{\Lambda} v^2 d\mu + 2 \sum_{0 \leq j < k \leq n-1} \int_{\Lambda} v v \circ T^{k-j} d\mu \\ &= n \int_{\Lambda} v^2 d\mu + 2 \sum_{r=1}^{n-1} (n-r) \int_{\Lambda} v v \circ T^r d\mu. \end{aligned}$$

Therefore

$$\frac{1}{\sqrt{n}} \int_{\Lambda} v_n^2 d\mu = \int_{\Lambda} v^2 d\mu + 2 \sum_{r=1}^{n-1} \int_{\Lambda} v v \circ T^r d\mu - (2/n) \sum_{r=1}^{n-1} r \int_{\Lambda} v v \circ T^r d\mu.$$

Taking the limit as $n \rightarrow \infty$ we conclude that

$$\sigma^2 = \lim_{n \rightarrow \infty} \int_{\Lambda} \left(\frac{1}{\sqrt{n}} v_n \right)^2 d\mu = \int_{\Lambda} v^2 d\mu + 2 \sum_{n=1}^{\infty} \int_{\Lambda} v v \circ T^n d\mu.$$

■

Corollary 4.6. $\sigma^2 = 0$ if and only if $v = \chi \circ T - \chi$ for some χ with $\|\chi\|_{\infty} < \infty$ (v is a coboundary).

Proof: First assume that $\sigma^2 = 0$. Since by Proposition 4.4 $\sigma^2 = \int_{\Lambda} m^2 d\mu$ this implies that $m = 0$ almost everywhere. The martingale approximation (3.1) in Theorem 3.9 then gives $v = \chi \circ T - \chi$. For the converse assume that $v = \chi \circ T - \chi$ which gives by Lemma 4.2 that $v_n = \chi \circ T^n - \chi$. Therefore $\|v_n\|_{\infty} \leq 2 \|\chi\|_{\infty}$ is bounded and we conclude that $\sigma^2 = (1/n) \lim_{n \rightarrow \infty} \int_{\Lambda} v_n^2 d\mu = 0$. ■

Now, recall that the *characteristic function* ϕ_Y of a random variable Y with values in \mathbb{R} is defined by $\phi_Y(t) = \mathbb{E}[e^{itY}]$ for all $t \in \mathbb{R}$. For a normally distributed random variable $Y \sim \mathcal{N}(0, \sigma^2)$ its characteristic function is $\phi_Y(t) = e^{t^2 \sigma^2 / 2}$.

We have the following classical result [24]:

Lemma 4.7. (Lévy's Continuity Theorem) *Let $Y, Y_n, n \in \mathbb{N}$ be random variables with corresponding characteristic functions $\phi_Y, \phi_{Y_n}, n \in \mathbb{N}$. Then $Y_n \rightarrow_d Y$ if and only if $\lim_{n \rightarrow \infty} \phi_{Y_n}(t) = \phi_Y(t)$ for all $t \in \mathbb{R}$.*

Proof of Theorem 4.1: Firstly let us note that it follows from Lemma 4.3 that the problem is reduced to proving that $\frac{1}{\sqrt{n}} m_n \rightarrow_d \mathcal{N}(0, \sigma^2)$ as $n \rightarrow \infty$. By Lévy's Continuity Theorem it is sufficient to prove that

$$\int_{\Lambda} e^{(it/\sqrt{n})m_n} d\mu \rightarrow e^{-t^2 \sigma^2 / 2} \quad \text{for all } t \in \mathbb{R}.$$

Consider the Taylor expansion $\log(z+1) = z - \frac{z^2}{2} + \frac{z^3}{3} - \dots$ for $|z| < 1$. Then

$$e^z = (1+z)e^{\frac{z^2}{2} - \frac{z^3}{3} + \dots} = (1+z)e^{\frac{z^2}{2} - r(z)}$$

where $|r(z)| \leq |z|^3/3$. Therefore we have

$$\exp\left(\frac{it}{\sqrt{n}} m \circ T^j\right) = \left(1 + \frac{it}{\sqrt{n}} m \circ T^j\right) \exp\left(-\frac{t^2}{2n} m^2 \circ T^j + r\left(\frac{it}{\sqrt{n}} m \circ T^j\right)\right).$$

It follows that

$$\begin{aligned}
\exp\left(\frac{it}{\sqrt{n}}m_n\right) &= \exp\left(\frac{it}{\sqrt{n}}\sum_{j=0}^{n-1}m \circ T^j\right) = \prod_{j=0}^{n-1}\exp\left(\frac{it}{\sqrt{n}}m \circ T^j\right) \\
&= \underbrace{\left[\prod_{j=0}^{n-1}\left(1 + \frac{it}{\sqrt{n}}m \circ T^j\right)\right]}_{T_n} \exp\left(-\frac{t^2}{2n}\sum_{j=0}^{n-1}m^2 \circ T^j + \sum_{j=0}^{n-1}r\left(\frac{it}{\sqrt{n}}m \circ T^j\right)\right) \\
&= T_n S_n.
\end{aligned} \tag{4.2}$$

Before concluding the proof of Theorem 4.1 we show the following:

Lemma 4.8. *Let S_n, T_n be defined by (4.2). The following hold true:*

- $\int_{\Lambda} T_n d\mu = 1$.
- $S_n \rightarrow \exp\left(-\frac{1}{2}t^2\sigma^2\right)$ as $n \rightarrow \infty$ a.e.
- Moreover, T_n and S_n are uniformly bounded.

Proof: We have

$$T_n = 1 + \frac{it}{\sqrt{n}}\sum_{j=0}^{n-1}m \circ T^j + \left(\frac{it}{\sqrt{n}}\right)^2 \sum_{\substack{j=0 \\ j \neq k}}^{n-1} m \circ T^j m \circ T^k + \dots + \left(\frac{it}{\sqrt{n}}\right)^n \prod_{j=0}^{n-1} m \circ T^j.$$

Since by Theorem 3.9 m is centered and using the multiplicative property for m , Proposition 3.11, we have that $\int_{\Lambda} T_n d\mu = 1$.

Furthermore,

$$\left|\sum_{j=0}^{n-1}r\left(\frac{it}{\sqrt{n}}m \circ T^j\right)\right| \leq \left|\frac{t}{\sqrt{n}}\right|^3 \sum_{j=0}^{n-1}|m^3 \circ T^j| \leq \left|\frac{t}{\sqrt{n}}\right|^3 \sum_{j=0}^{n-1}\|m^3\|_{\infty} = \frac{|t^3|}{\sqrt{n}}\|m^3\|_{\infty}.$$

Applying Birkhoff's Ergodic Theorem to the partial sums $n^{-1} \sum_{j=0}^{n-1} m^2 \circ T^j$ and taking exponentials we obtain that

$$S_n \rightarrow \exp \left(-\frac{1}{2} t^2 \int_{\Lambda} m^2 d\mu \right) = \exp \left(-\frac{1}{2} t^2 \sigma^2 \right) \quad \text{a.e.}$$

Next we have that

$$\begin{aligned} |T_n| &= \prod_{j=0}^{n-1} \left| 1 + \frac{it}{\sqrt{n}} m \circ T^j \right| = \prod_{j=0}^{n-1} \sqrt{1 + \frac{t^2}{n} m^2 \circ T^j} \\ &\leq \prod_{j=0}^{n-1} \sqrt{1 + \frac{t^2}{n} \|m^2 \circ T^j\|_{\infty}} \leq \sqrt{\left(1 + \frac{t^2}{n} \|m^2\|_{\infty} \right)^n} \\ &\xrightarrow{n \rightarrow \infty} \exp(t^2 \|m^2\|_{\infty} / 2) \end{aligned}$$

which proves that T_n is uniformly bounded. Moreover,

$$\|S_n\|_{\infty} \leq \exp(|t^3| \|m^3\|_{\infty}) \quad \text{for all } n \in \mathbb{N}$$

gives that S_n is uniformly bounded¹. ■

Recall Lebesgue's Dominated Convergence Theorem [74].

Lemma 4.9. (Dominated Convergence Theorem) *Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of measurable functions on $(\Lambda, \mathcal{F}, \mu)$. Suppose there exists an integrable function g such that $|f_n| \leq g$ for all $n \in \mathbb{N}$. If $f_n \rightarrow f$ almost everywhere then f is integrable and*

$$\lim_{n \rightarrow \infty} \int_{\Lambda} f_n d\mu = \int_{\Lambda} \lim_{n \rightarrow \infty} f_n d\mu = \int_{\Lambda} f d\mu.$$

¹Alternatively, one can note that $\|S_n\|_{\infty} = \left\| \exp \left(\frac{it}{\sqrt{n}} m_n \right) / T_n \right\|_{\infty} \leq 1$.

Proof of Theorem 4.1, continued: Now Lemma 4.8 gives that $T_n S_n$ is uniformly bounded. Therefore we can apply Lemma 4.9. We have

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \int_{\Lambda} \exp\left(\frac{it}{\sqrt{n}} m_n\right) d\mu = \lim_{n \rightarrow \infty} \int_{\Lambda} T_n S_n d\mu \\
&= \lim_{n \rightarrow \infty} \int_{\Lambda} \left[T_n \exp\left(-\frac{1}{2} t^2 \sigma^2\right) + T_n \left(S_n - \exp\left(-\frac{1}{2} t^2 \sigma^2\right) \right) \right] d\mu \\
&= \exp\left(-\frac{1}{2} t^2 \sigma^2\right) \lim_{n \rightarrow \infty} \int_{\Lambda} T_n d\mu + \int_{\Lambda} \lim_{n \rightarrow \infty} T_n \left(S_n - \exp\left(-\frac{1}{2} t^2 \sigma^2\right) \right) d\mu \\
&= \exp\left(-\frac{1}{2} t^2 \sigma^2\right),
\end{aligned}$$

where the third equality follows from the Dominated Convergence Theorem and the last equality follows from Lemma 4.8. Thus by Lévy's Continuity Theorem we have that

$$\frac{1}{\sqrt{n}} m_n \rightarrow_d \mathcal{N}(0, \sigma^2)$$

and therefore by Lemma 4.3 we conclude that

$$\frac{1}{\sqrt{n}} v_n \rightarrow_d \mathcal{N}(0, \sigma^2).$$

■

4.3 Martingale Theory

We have mentioned in chapter 3 that the decomposition of v_n into a martingale part m_n plus a telescoping sum of random variables gives an alternative way to study limit theorems by exploiting results from martingale theory. We now gather together some standard concepts from probability theory and martingale theory in particular and then proceed to obtain a martingale Central Limit Theorem.

In what follows $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. We begin with a review of conditional expectations [89].

Theorem 4.10. (Fundamental Theorem, Kolmogorov [49]) *Let Z be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ -algebra on Ω . There exists a random variable Y such that*

- (i) $\mathbb{E}|Y| < \infty$,
- (ii) Y is \mathcal{G} -measurable,
- (iii) $\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A]$ for all $A \in \mathcal{G}$.

If \hat{Y} is another random variable satisfying (i)-(iii) then $\hat{Y} = Y$, \mathbb{P} -a.s. We call Y a version of $\mathbb{E}[Z | \mathcal{G}]$ and write $Y = \mathbb{E}[Z | \mathcal{G}]$, \mathbb{P} -a.s.

This tackles existence and uniqueness (up to equality almost surely) of conditional expectations. We list below for easy reference some basic properties of conditional expectations [89].

Proposition 4.11. *Let $Z, X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ sub- σ -algebras on Ω .*

- CE(a) $\mathbb{E}[cZ + X | \mathcal{G}] = c\mathbb{E}[Z | \mathcal{G}] + \mathbb{E}[X | \mathcal{G}]$ for all $c \in \mathbb{R}$.*
- CE(b) If Z is \mathcal{G} -measurable then $\mathbb{E}[Z | \mathcal{G}] = Z$ \mathbb{P} -a.s.*
- CE(c) $\mathbb{E}[\mathbb{E}[Z | \mathcal{G}]] = \mathbb{E}[Z]$ \mathbb{P} -a.s.*
- CE(d) If X is \mathcal{G} -measurable and bounded then $\mathbb{E}[XZ | \mathcal{G}] = X\mathbb{E}[Z | \mathcal{G}]$ \mathbb{P} -a.s.*
- CE(e) $\|\mathbb{E}[Z | \mathcal{G}]\|_p \leq \|Z\|_p$ for $p \geq 1$.*

We can now define a martingale. Let $\{\mathcal{F}_n\}_{n \geq 0}$ be a family of σ -algebras on Ω such that $\mathcal{F}_n \subset \mathcal{F}$ for all $n \in \mathbb{N}$. Recall that $\{\mathcal{F}_n\}_{n \geq 0}$ is a *filtration* on \mathcal{F} if $\mathcal{F}_m \subset \mathcal{F}_n$ for $m \leq n$.

Definition 4.12. *A stochastic process $\{Z_n\}_{n \geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n \geq 0}$ if*

- (a) $\mathbb{E}[|Z_n|] < \infty$ for all $n > 0$,
- (b) Z_n is \mathcal{F}_n -adapted, i.e. Z_n is \mathcal{F}_n -measurable for all $n > 0$,
- (c) $\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = Z_n$ a.s for all $n > 0$.

We call $\{Z_n\}_{n \geq 0}$ a martingale difference sequence, (m.d.s.), with respect to the filtration $\{\mathcal{F}_n\}_{n \geq 0}$ if it satisfies (a), (b), and it has the property,

$$(c') \quad \mathbb{E}[Z_{n+1} | \mathcal{F}_n] = 0 \text{ a.s for all } n > 0.$$

It is easy to see from Definition 4.12 that there is a strong link between a martingale difference sequence and a martingale.

Proposition 4.13. *The following hold true:*

- i) If $\{X_n\}_{n \geq 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n \geq 0}$ then $\{Y_n\}_{n \geq 0}$ defined by $Y_0 = X_0$ and $Y_n = X_n - X_{n-1}$ for $n \geq 1$ is a m.d.s. with respect to $\{\mathcal{F}_n\}_{n \geq 0}$.
- ii) If $\{Y_n\}_{n \geq 0}$ is a m.d.s. with respect to $\{\mathcal{F}_n\}_{n \geq 0}$ then $\{X_n\}_{n \geq 0}$ defined by $X_n = \sum_{i=0}^n Y_i$ is a martingale with respect to $\{\mathcal{F}_n\}_{n \geq 0}$.

Proof: Clearly we have that integrability and \mathcal{F}_n -adaptedness hold in both cases. For i) observe that if $\{X_n\}_{n \geq 0}$ is a martingale then for all $n \geq 0$

$$\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - X_n = X_n - X_n = 0.$$

Note that the second equality above follows from Definition 4.12 (b) and properties CE(a), CE(b) of Proposition 4.11 and the third equality follows from Definition 4.12 (c). If $\{Y_n\}_{n \geq 0}$ is a m.d.s. then

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[Y_{n+1} + \sum_{i=0}^n Y_i | \mathcal{F}_n] = \mathbb{E}[Y_{n+1} | \mathcal{F}_n] + \sum_{i=0}^n Y_i = X_n$$

where the second equality follows from Definition 4.12 (b) and properties CE(a), CE(b) of Proposition 4.11 and the third equality from Definition 4.12 (c'). ■

The definition of a martingale difference sequence has a dual definition for stochastic processes defined with respect to a decreasing sequence of σ -algebras. This is the concept of reverse martingales.

Definition 4.14. Let $\{\mathcal{F}_n\}_{n \geq 1}$ be a reverse filtration i.e. $\mathcal{F}_{n+1} \subset \mathcal{F}_n$ for all $n \geq 1$. A stochastic process $\{Z_n\}_{n \geq 1}$ is called a reverse martingale difference sequence with respect to $\{\mathcal{F}_n\}_{n \geq 1}$ if

- (i) $\mathbb{E}[|Z_n|] < \infty$ for all $n \geq 1$,
- (ii) Z_n is \mathcal{F}_n -adapted, i.e. Z_n is \mathcal{F}_n -measurable for all $n \geq 1$,
- (iii) $\mathbb{E}[Z_n | \mathcal{F}_{n+1}] = 0$ a.s for all $n \geq 1$.

Now, we have seen in the previous section that the multiplicativity property for m in Proposition 3.11 is central to the proof of the Central Limit Theorem compensating in a way for the lack of independence. We note below a relation between martingale difference sequences and Proposition 3.11.

Proposition 4.15. A martingale difference sequence $\{Y_n, \mathcal{F}_n\}_{n \geq 1}$ is orthogonal.

Proof: Let $j < i$. Then by Proposition 4.11, properties CE(c) and CE(d), we have

$$\mathbb{E}[Y_i Y_j] = \mathbb{E}[\mathbb{E}[Y_i Y_j | \mathcal{F}_{i-1}]] = \mathbb{E}[Y_j \mathbb{E}[Y_i | \mathcal{F}_{i-1}]] = 0.$$

More generally let $j_1 < j_2 < \dots < j_i$, $j_i \in \mathbb{N}$ for all $i \in \mathbb{N}$. We have

$$\mathbb{E} \left[\prod_{i=0}^i Y_{j_i} \right] = \mathbb{E} \left[\mathbb{E} \left[Y_i \prod_{k=0}^{i-1} Y_{j_k} \mid \mathcal{F}_{i-1} \right] \right] = \mathbb{E} \left[\prod_{k=0}^{i-1} Y_{j_k} (\mathbb{E}[Y_i | \mathcal{F}_{i-1}]) \right] = 0.$$

■

4.4 Central Limit Theorem via Martingales

Let us begin by recalling the following property of a measure-preserving map:

Lemma 4.16. *Let $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subset \mathcal{F}$ a σ -algebra. If T is measure-preserving then*

$$\mathbb{E}[Y \circ T \mid T^{-1}\mathcal{G}] = \mathbb{E}[Y \mid \mathcal{G}] \circ T.$$

Proof: Firstly note that

$$\{x : \mathbb{E}[Y \mid \mathcal{G}](x) \leq c\} \in \mathcal{G} \quad \text{for all } c \in \mathbb{R}.$$

Therefore

$$\{x : \mathbb{E}[Y \mid \mathcal{G}](Tx) \leq c\} = T^{-1}\{x : \mathbb{E}[Y \mid \mathcal{G}](x) \leq c\} \in T^{-1}\mathcal{G}$$

and we have that $\mathbb{E}[Y \mid \mathcal{G}] \circ T$ is $T^{-1}\mathcal{G}$ -measurable. Next observe that for any $A \in \mathcal{G}$

$$\begin{aligned} \int_{T^{-1}A} \mathbb{E}[Y \mid \mathcal{G}] \circ T \, d\mu &= \int \mathbf{1}_{T^{-1}A} \mathbb{E}[Y \mid \mathcal{G}] \circ T \, d\mu = \int (\mathbf{1}_A \circ T) \mathbb{E}[Y \mid \mathcal{G}] \circ T \, d\mu \\ &= \int \mathbf{1}_A \mathbb{E}[Y \mid \mathcal{G}] \, d\mu = \int_A \mathbb{E}[Y \mid \mathcal{G}] \, d\mu = \int_A Y \, d\mu, \end{aligned}$$

where the third equality follows from T -invariance for μ . Similarly

$$\begin{aligned} \int_{T^{-1}A} \mathbb{E}[Y \circ T \mid T^{-1}\mathcal{G}] \, d\mu &= \int_{T^{-1}A} Y \circ T \, d\mu = \int \mathbf{1}_{T^{-1}A} Y \circ T \, d\mu \\ &= \int (\mathbf{1}_A \circ T) Y \circ T \, d\mu = \int \mathbf{1}_A Y \, d\mu = \int_A Y \, d\mu. \end{aligned}$$

Thus we have

$$\int_{T^{-1}A} \mathbb{E}[Y \circ T \mid T^{-1}\mathcal{G}] \, d\mu = \int_{T^{-1}A} \mathbb{E}[Y \mid \mathcal{G}] \circ T \, d\mu$$

for all $A \in \mathcal{G}$. Lastly, the integrability is immediate since $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Hence we conclude the claim by Theorem 4.10. ■

We have seen in Proposition 3.3 that $\mathcal{P}\mathcal{U}v = v$. We also have the following:

Proposition 4.17. *If $v \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ then $\mathcal{U}\mathcal{P}v = \mathbb{E}[v \mid T^{-1}\mathcal{F}]$.*

Proof: Note that since $\mathbf{1}_{T^{-1}A}(x) = 1$ if and only if $x \in T^{-1}A$ if and only if $T(x) \in A$ if and only if $\mathbf{1}_A T(x) = 1$, it holds true that $\mathbf{1}_{T^{-1}A} = \mathcal{U}\mathbf{1}_A$. Let $A \in T^{-1}\mathcal{F}$, that is $A = T^{-1}B$ for some $B \in \mathcal{F}$. Using the adjoint-property of the transfer operator \mathcal{P} , Definition 3.2, we then have

$$\begin{aligned} \int_A (\mathcal{U}\mathcal{P}v) d\mu &= \int_{T^{-1}B} (\mathcal{U}\mathcal{P}v) d\mu = \int (\mathcal{U}\mathcal{P}v) \mathbf{1}_{T^{-1}B} d\mu = \int (\mathcal{U}\mathcal{P}v) \mathcal{U}\mathbf{1}_B d\mu \\ &= \int (\mathcal{P}v) \mathbf{1}_B d\mu = \int v (\mathcal{U}\mathbf{1}_B) d\mu = \int v \mathbf{1}_{T^{-1}B} d\mu = \int_A v d\mu. \end{aligned}$$

Since $\mathcal{U}\mathcal{P}v = \mathcal{P}v \circ T$ is $T^{-1}\mathcal{F}$ -measurable it only remains to check the integrability condition. Observe that by Proposition 3.3 we have

$$\mathbb{E} |\mathcal{U}\mathcal{P}v| = \|\mathcal{U}\mathcal{P}v\|_1 = \|\mathcal{P}v\|_1 \leq \|v\|_1 = \mathbb{E} |v| \leq (\mathbb{E} |v|^2)^{1/2} = \|v\|_2 < \infty$$

since $v \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Thus we conclude by Proposition 4.10 that

$$\mathcal{U}\mathcal{P}v = \mathbb{E} [v \mid T^{-1}\mathcal{F}].$$

■

Proposition 4.18. *Assume that T satisfies the General Framework (H) and let $(\Lambda, \mathcal{F}, \mu)$ be the underlying probability space. Let m be the function associated to v , where is Hölder continuous and mean zero, as in Theorem 3.9. Then $\{m \circ T^j\}_{j \geq 0}$ is a sequence of reverse martingale differences with respect to the reverse filtration $\{T^{-j}\mathcal{F}\}_{j \geq 0}$.*

Proof: The integrability condition is trivial since $m \in \mathcal{L}^2(\Lambda, \mathcal{F}, \mu)$. Next, it is clear that $m \circ T^j$ is $T^{-j}\mathcal{F}$ -measurable for $j \geq 0$. However, since

$$T^{-1}\mathcal{F} = \{T^{-1}(A) \mid A \in \mathcal{F}\} \subset \mathcal{F},$$

$$T^{-2}\mathcal{F} = \{T^{-2}(A) \mid A \in \mathcal{F}\} = \{T^{-1}(\hat{A}) \mid \hat{A} \in T^{-1}\mathcal{F}\} \subset T^{-1}\mathcal{F}, \dots,$$

we obtain a decreasing sequence of sub- σ -algebras $\{T^{-j}\mathcal{F}\}_{j \geq 0}$ of \mathcal{F} satisfying $T^{-n}\mathcal{F} \supset T^{-m}\mathcal{F}$ for $m \geq n > 0$ i.e. a reverse filtration. Lastly let us note that by Lemma 4.16, Proposition 4.17, and since, by Theorem 3.9, $m \in \text{Ker}(\mathcal{P})$ we have that for all $j \geq 0$

$$\mathbb{E} [m \circ T^j \mid T^{-(j+1)}\mathcal{F}] = \mathbb{E} [m \mid T^{-1}\mathcal{F}] \circ T^j = \mathcal{UP}m \circ T^j = 0.$$

■

Remark 4.19. *For a basic paradigm for the CLT one can consider the doubling map where we take $([0, 1], \mathcal{B}([0, 1]), \lambda)$ with $\mathcal{B}([0, 1])$ being the Borel σ -algebra on the unit interval and λ is Lebesgue measure. Then as we have seen in Theorem 3.6 the dynamics of the doubling map satisfies the General Framework (H) and therefore by Theorem 3.9 a martingale approximation exists.*

As we have seen in Proposition 4.18, $\{T^{-j}\mathcal{F}\}_{j \geq 0}$ is a reverse filtration, it goes in the wrong direction. In order to call into use the preceding martingale theory methods we need to have a filtration. To this end we pass from the non-invertible map T to an invertible map \tilde{T} with similar dynamical properties. This is achieved using *natural extensions* [71, 72], also [64, 66].

Proposition 4.20. (Natural extension) *Let $(\Lambda, \mathcal{F}, \mu)$ be a measure space and $T : (\Lambda, \mathcal{F}, \mu) \rightarrow (\Lambda, \mathcal{F}, \mu)$ a possibly noninvertible measure-preserving transformation. There exists an invertible measure-preserving transformation $\tilde{T} : (\tilde{\Lambda}, \tilde{\mathcal{F}}, \tilde{\mu}) \rightarrow (\tilde{\Lambda}, \tilde{\mathcal{F}}, \tilde{\mu})$ and a map $\pi : \tilde{\Lambda} \rightarrow \Lambda$ with the following properties:*

(NE1) \tilde{T} is an extension of T in the sense that $\pi \circ \tilde{T} = T \circ \pi$, (see figure 4.1).

(NE2) $\pi_*\tilde{\mu} = \mu$ i.e. $\tilde{\mu}(\pi^{-1}A) = \mu(A)$.

(NE3) $\tilde{\mu}$ is ergodic w.r.t. \tilde{T} if and only if μ is ergodic w.r.t. T .

(NE4) The sub- σ -algebra $\tilde{\mathcal{F}} \subset \hat{\mathcal{F}}$ generated by sets $\{\pi^{-1}A \mid A \in \mathcal{F}\}$ satisfies $\dots \subset \tilde{T}^{-1}\tilde{\mathcal{F}} \subset \tilde{\mathcal{F}} \subset \tilde{T}\tilde{\mathcal{F}} \subset \tilde{T}^2\tilde{\mathcal{F}} \subset \dots$

$$\begin{array}{ccc}
\tilde{\Lambda} & \xrightarrow{\tilde{T}} & \tilde{\Lambda} \\
\pi \downarrow & & \downarrow \pi \\
\Lambda & \xrightarrow{T} & \Lambda
\end{array}$$

Figure 4.1: $\pi \circ \tilde{T} = T \circ \pi$

Remark 4.21. The map π is called a homomorphism or a factor map [64]. The map \tilde{T} is called the natural extension of T and it is unique up to isomorphism e.g. [66].

Lemma 4.22. Let $\tilde{T} : (\tilde{\Lambda}, \hat{\mathcal{F}}, \tilde{\mu}) \rightarrow (\tilde{\Lambda}, \hat{\mathcal{F}}, \tilde{\mu})$ be the natural extension of $T : (\Lambda, \mathcal{F}, \mu) \rightarrow (\Lambda, \mathcal{F}, \mu)$. Denote the lifted observable by

$$\tilde{m} = m \circ \pi : \tilde{\Lambda} \rightarrow \mathbb{R}.$$

Then $\int_{\tilde{\Lambda}} \tilde{m} d\tilde{\mu} = \int_{\tilde{\Lambda}} \tilde{m} \circ \tilde{T}^j d\tilde{\mu} = 0$ and $\int_{\tilde{\Lambda}} \tilde{m}^2 d\tilde{\mu} = \sigma^2$. Furthermore, $\|m\|_p < \infty$ implies $\|\tilde{m}\|_p < \infty$.

Proof: For $i = 1, 2$ we have

$$\int_{\tilde{\Lambda}} \tilde{m}^i d\tilde{\mu} = \int_{\tilde{\Lambda}} m^i \circ \pi d\tilde{\mu} = \int_{\Lambda} m^i d(\pi_* \tilde{\mu}) = \int_{\Lambda} m^i d\mu.$$

In particular, $\int_{\tilde{\Lambda}} \tilde{m} d\tilde{\mu} = 0$ and $\int_{\tilde{\Lambda}} \tilde{m}^2 d\tilde{\mu} = \sigma^2$. Also, since

$$\tilde{m} \circ \tilde{T}^j = m \circ \pi \circ \tilde{T}^j = m \circ T^j \circ \pi$$

we have

$$\begin{aligned}
\int_{\tilde{\Lambda}} \tilde{m} \circ \tilde{T}^j d\tilde{\mu} &= \int_{\tilde{\Lambda}} m \circ T^j \circ \pi d\tilde{\mu} = \int_{\Lambda} m \circ T^j d(\pi_* \tilde{\mu}) \\
&= \int_{\Lambda} m \circ T^j d\mu = \int_{\Lambda} m d\mu = 0.
\end{aligned}$$

If $\|m\|_p < \infty$ then

$$\begin{aligned}\|\tilde{m}\|_p^p &= \int_{\tilde{\Lambda}} |\tilde{m}|^p d\tilde{\mu} = \int_{\tilde{\Lambda}} |m \circ \pi|^p d\tilde{\mu} = \int_{\tilde{\Lambda}} |m|^p \circ \pi d\tilde{\mu} \\ &= \int_{\Lambda} |m|^p d(\pi_*\tilde{\mu}) = \int_{\Lambda} |m|^p d\mu = \|m\|_p^p.\end{aligned}$$

■

Proposition 4.23. *Define $\tilde{m}_n^- = \sum_{j=1}^n \tilde{m} \circ \tilde{T}^{-j}$. Then \tilde{m}_n^- is a martingale with respect to $\{\tilde{T}^j \tilde{\mathcal{F}}\}_j$.*

Proof: It is clear that $\tilde{m} \circ \tilde{T}^{-j}$ is $\tilde{T}^j \tilde{\mathcal{F}}$ -measurable for all $j \geq 0$. Also, by Proposition 4.20 (NE4) we have that $\{\tilde{T}^j \tilde{\mathcal{F}}\}_j$ is a filtration. Since π and \tilde{T} are measure-preserving the integrability condition is also satisfied:

$$\mathbb{E} \left| \tilde{m} \circ \tilde{T}^{-j} \right| = \left\| \tilde{m} \circ \tilde{T}^{-j} \right\|_1 = \|\tilde{m}\|_1 = \|m\|_1 \leq \|m\|_2 < \infty.$$

Moreover, by Lemma 4.16 we have

$$\begin{aligned}\mathbb{E} \left[\tilde{m} \circ \tilde{T}^{-j} \mid \tilde{T}^{j-1} \tilde{\mathcal{F}} \right] &= \mathbb{E} \left[\tilde{m} \mid \tilde{T}^{-1} \tilde{\mathcal{F}} \right] \circ \tilde{T}^{-j} \\ &= \mathbb{E} \left[m \circ \pi \mid \tilde{T}^{-1} \pi^{-1} \mathcal{F} \right] \circ \tilde{T}^{-j} = \mathbb{E} \left[m \circ \pi \mid \pi^{-1} T^{-1} \mathcal{F} \right] \circ \tilde{T}^{-j} \\ &= \mathbb{E} \left[m \mid T^{-1} \mathcal{F} \right] \circ \pi \circ \tilde{T}^{-j} = 0,\end{aligned}$$

where the last equality follows from $m \in \text{Ker}(\mathcal{P})$ (Theorem 3.9) and Proposition 4.17. It follows that $\{\tilde{m} \circ \tilde{T}^{-j}\}_{j \geq 0}$ is a martingale difference sequence with respect to $\{\tilde{T}^j \tilde{\mathcal{F}}\}_j$. Hence we conclude by Proposition 4.13 that \tilde{m}_n^- is a martingale with respect to the filtration $\{\tilde{T}^j \tilde{\mathcal{F}}\}_j$. ■

Theorem 4.24. (Ergodic CLT) *Let $Y \in \mathcal{L}^2(\Lambda, \mathcal{F}, \mu)$ where $(\Lambda, \mathcal{F}, \mu)$ is a probability space and $\{\mathcal{G}_i\}_i$ a filtration of \mathcal{F} . Let $T : \Lambda \rightarrow \Lambda$ be an ergodic, measure-preserving transformation and assume that $\mathbb{E}[Y^2] = \sigma^2$ is positive and finite. If the stationary, ergodic sequence $\{Y \circ T^i\}_i$ is a family of martingale differences i.e. for all $i \geq 0$ we have that $Y \circ T^i$ is \mathcal{G}_i -measurable and*

$$\mathbb{E}[Y \circ T^n \mid \mathcal{G}_{n-1}] = 0,$$

then

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} Y \circ T^i \rightarrow_d \mathcal{N}(0, \sigma^2) .$$

Remark 4.25. *The Central Limit Theorem for stationary and ergodic martingale differences, Theorem 4.24 above, was established, independently, by Billingsley [7] and Ibragimov [40]. Brown [15] and Dvoretzky [26] have also obtained analogous Central Limit Theorems.*

We conclude the chapter with the martingale theory proof of the Central Limit Theorem.

Proof of Theorem 4.1 (Martingale version): Since we have by Proposition 4.20 (NE3) that ergodicity is preserved under natural extensions and using Proposition 4.23 we obtain that \tilde{m}_n^- is a stationary ergodic martingale. It follows from Theorem 4.24 that

$$\frac{1}{\sqrt{n}} \tilde{m}_n^- \rightarrow_d \mathcal{N}(0, \sigma^2) .$$

By Proposition 4.20 we have

$$\{m \circ T^j\}_{j \geq 0} =_d \{\tilde{m} \circ \tilde{T}^j\}_{j \geq 0}$$

and we therefore obtain that

$$m_n =_d \sum_{j=0}^{n-1} \tilde{m} \circ \tilde{T}^j = \tilde{m}_n .$$

Moreover,

$$\tilde{m}_n^- \circ \tilde{T}^n = \left(\sum_{j=1}^n \tilde{m} \circ \tilde{T}^{-j} \right) \circ \tilde{T}^n = \sum_{j=1}^n \tilde{m} \circ \tilde{T}^{-j+n} = \sum_{j=0}^{n-1} \tilde{m} \circ \tilde{T}^j = \tilde{m}_n .$$

Hence we have that $m_n =_d \tilde{m}_n =_d \tilde{m}_n \circ \tilde{T}^{-n} = \tilde{m}_n^-$. Thus we deduce the Central Limit Theorem for m_n ,

$$\frac{1}{\sqrt{n}} m_n \rightarrow_d \mathcal{N}(0, \sigma^2) .$$

We conclude by Lemma 4.3 that

$$\frac{1}{\sqrt{n}}v_n \rightarrow_d \mathcal{N}(0, \sigma^2) .$$

■

Chapter 5

Rate of Convergence in the Central Limit Theorem

5.1 Introduction

Having proved the Central Limit Theorem it is natural to raise the question of what is the related speed of convergence. Let us recall the classical corresponding result for independent, identically distributed random variables: Consider a sequence $\{X_i\}_{i \in \mathbb{N}}$ of i.i.d. r.v.'s such that $\mathbb{E}[X_1] = 0$ and $\sigma^2 = \mathbb{E}[X_1^2]$ and let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n}{\sigma\sqrt{n}} \rightarrow_d \mathcal{N}(0, 1).$$

If $\{X_i\}_{i \in \mathbb{N}}$ has bounded third order moments, $\mathbb{E}[|X_1|^3] = \gamma < \infty$, then the celebrated Berry-Esseen Theorem [5, 27] gives the optimal rate of convergence in the Central Limit Theorem: there exists a universal constant $C > 0$ such that

$$\sup_t \left| \mathbb{P}[S_n/\sqrt{n} \leq \sigma t] - \Phi(t) \right| \leq C \frac{\gamma}{\sigma^3 \sqrt{n}}$$

where $\Phi(t)$ is the distribution function for a random variable $Y \sim \mathcal{N}(0, 1)$ i.e. its distribution function is

$$\Phi(t) = \mathbb{P}[Y \leq t] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2}} ds.$$

In the previous chapter we have seen that under the General Framework **(H)** the Central Limit Theorem for the ergodic, stationary stochastic process $\{v \circ T^i\}_{i \geq 0}$ holds true:

$$\frac{v_n}{\sigma\sqrt{n}} \rightarrow_d \mathcal{N}(0, 1)$$

where $v_n = \sum_{i=0}^{n-1} v \circ T^i$ and $\sigma^2 = \int m^2 d\mu$. We now ask for uniform bounds in the distance between the distribution function $\mu(v_n/\sqrt{n} \leq \sigma t)$ and Φ :

$$D(v_n) = \sup_t |\mu(v_n/\sqrt{n} \leq \sigma t) - \Phi(t)|.$$

5.2 Rate of Convergence for Martingale Approximations

We prove the following:

Theorem 5.1. (Rate of Convergence in the CLT) *Assume that T satisfies the General Framework **(H)** and is Lipschitz continuous. Let $v \in C^\alpha(\Lambda)$ with $\int_\Lambda v d\mu = 0$. Then for all $\delta > 0$ there exists a constant $C > 0$ such that*

$$D(v_n) \leq C n^{-1/4+\delta}.$$

Remark 5.2. *We have seen that for a Hölder observable v and the doubling map T , χ is Hölder continuous. We now require smoothness throughout the martingale approximation (3.1) for the doubling map, namely for $\chi \circ T$ as well. Consider the unit circle $\mathbb{S}^1 := [0, 1]/\sim$, where \sim indicates that points 0 and 1 are identified. The doubling map in multiplicative notation is $T(w) = w^2$ where $w = e^{2\pi i \theta}$. We obtain Lipschitzness of the doubling map, with Lipschitz*

constant $\text{Lip}(T) = 2$, from the (normalized) arc length metric:

$$d(w_1, w_2) = \frac{1}{2\pi} \min \{ |\theta_1 - \theta_2|, 1 - |\theta_1 - \theta_2| \} .$$

Note that if χ is Hölder continuous and T is Lipschitz we then have that $\chi \circ T$ is Hölder.

Remark 5.3. We aim to set limits on the largest deviation of $\mu(v_n/\sqrt{n} \leq \sigma t)$ from Φ . To this end let us first recall that by Proposition 4.23 the process $\{\tilde{m} \circ \tilde{T}^{-j}\}_j$, obtained by passing to the natural extension using Proposition 4.20, is a martingale difference sequence with respect to the filtration $\{\tilde{T}^j \tilde{\mathcal{F}}\}_j$. Hence existing results in the literature for the rate of convergence in the CLT for martingales can be called upon and applied to the current setting. Specifically, we use a result by Mourrat in [58] in which the author also summarises and builds on previous results by [11, 36, 37, 43].

Denote the conditional variance by

$$V_n = \sum_{i=1}^n \mathbb{E}[\tilde{m}^2 \circ \tilde{T}^{-i} \mid \tilde{T}^{i-1} \tilde{\mathcal{F}}] . \quad (5.1)$$

We use the result on the rate of convergence in the Central Limit Theorem for a martingale difference sequence $\{\tilde{m} \circ \tilde{T}^{-j}\}_j$ given in [58], Theorem 1.5, which is simplified in our context to the following:

Theorem 5.4. Let $p \in [1, \infty)$. There exists a constant $C > 0$ (depending only on p and $\|m\|_\infty$) such that for any $n \geq 2$,

$$D(\tilde{m}_n^-) \leq C \left[\frac{n \log n}{(n\sigma^2)^{3/2}} + \left(\left\| \frac{V_n}{n\sigma^2} - 1 \right\|_p^p + \frac{1}{(n\sigma^2)^p} \right)^{1/(2p+1)} \right] .$$

Let us begin by recalling Burkholder's inequality [17].

Proposition 5.5. (Burkholder's inequality) *Let $p \geq 2$ and assume that $S_n = \sum_{i=1}^n X_i$ is a martingale with $\|X_i\|_p < \infty$ for all $i \geq 1$. There exists a real constant $C > 0$ such that $\|\max_{j \leq n} |S_j|\|_p \leq Cn^{1/2} \max_{j \leq n} \|X_j\|_p$ for all $n \geq 1$.*

We deduce from Burkholder's inequality the following:

Corollary 5.6. *Assume that $T : \Lambda \rightarrow \Lambda$ satisfies the General Framework (H) and v is Hölder continuous and $\int_{\Lambda} v d\mu = 0$. Then for all p there exists a real constant $C > 0$ such that $\|v_n\|_p \leq Cn^{1/2}$ for all $n \geq 1$.*

Proof: First note that it follows from the General Framework (H), and therefore from Theorem 3.9, that there exists $\chi \in \mathcal{C}^\alpha(\Lambda)$ such that we have the decomposition $v_n = m_n + \chi \circ T^n - \chi$. By the properties of the natural extension (Proposition 4.20) and by Proposition 4.23 there exists a martingale process \tilde{m}_n^- where $\tilde{m}_n^- \circ \tilde{T}^n = \tilde{m}_n^- = m_n \circ \pi$. It follows from Proposition 5.5 that there exists a real constant $C > 0$ such that $\|\max_{j \leq n} |\tilde{m}_j^-|\|_p \leq Cn^{1/2}$ for all $n \geq 1$. Hence we have that there exist real positive constants C, C_1 such that

$$\begin{aligned} \|v_n\|_p &\leq \|m_n\|_p + \|\chi \circ T^n\|_p + \|\chi\|_p \leq \|\tilde{m}_n^-\|_p + 2\|\chi\|_p \\ &\leq C_1 n^{1/2} + 2\|\chi\|_p n^{1/2} \leq Cn^{1/2} \end{aligned}$$

for all $n \geq 1$. ■

Proposition 5.7. *Let $p \geq 2$ and assume that $T : \Lambda \rightarrow \Lambda$ satisfies the General Framework (H) and is Lipschitz continuous. Assume that v is Hölder continuous and that $\int_{\Lambda} v d\mu = 0$. There exists $C > 0$ such that*

$$\|(n\sigma^2)^{-1}V_n - 1\|_p \leq Cn^{-1/2}.$$

Proof: Define

$$\begin{aligned} \hat{v} &= \mathbb{E}[m^2 | T^{-1}\mathcal{F}] - \int (\mathbb{E}[m^2 | T^{-1}\mathcal{F}]) d\mu \\ &= \mathcal{UP}m^2 - \sigma^2. \end{aligned} \tag{5.2}$$

Observe that \hat{v} is mean-zero and that since m is Hölder continuous we have that m^2 is also Hölder. It follows by the General Framework **(H)** that $\mathcal{P}m^2$ is Hölder as well and therefore $\mathcal{UP}m^2 = \mathcal{P}m^2 \circ T$ is Hölder since T is, by assumption of Theorem 5.1, Lipschitz continuous. Using the properties of the natural extension, Proposition 4.20, we obtain that

$$\mathbb{E} \left[\tilde{m}^2 \circ \tilde{T}^{-i} \mid \tilde{T}^{i-1} \tilde{\mathcal{F}} \right] = \mathbb{E} \left[m^2 \mid T^{-1} \mathcal{F} \right] \circ \pi \circ \tilde{T}^{-i}. \quad (5.3)$$

We have

$$\begin{aligned} V_n &= \sum_{i=1}^n \mathbb{E} \left[\tilde{m}^2 \circ \tilde{T}^{-i} \mid \tilde{T}^{i-1} \tilde{\mathcal{F}} \right] \stackrel{(5.3)}{=} \sum_{i=1}^n \mathbb{E} \left[m^2 \mid T^{-1} \mathcal{F} \right] \circ \pi \circ \tilde{T}^{-i} \\ &= {}_d \sum_{i=1}^n \mathbb{E} \left[m^2 \mid T^{-1} \mathcal{F} \right] \circ T^i. \end{aligned}$$

Hence

$$V_n - n\sigma^2 \stackrel{(5.2)}{=} {}_d \sum_{i=1}^n (\mathcal{UP}m^2 - \sigma^2) \circ T^i = \sum_{i=1}^n \hat{v} \circ T^i = \hat{v}_n.$$

Since \hat{v} is mean-zero and Hölder continuous it follows from Corollary 5.6 that there exists a real constants $C_1, C > 0$ such that

$$\| (n\sigma^2)^{-1} V_n - 1 \|_p = (n\sigma^2)^{-1} \|\hat{v}_n\|_p \leq C_1 \sqrt{n} (n\sigma^2)^{-1} = Cn^{-1/2}$$

for all $n \geq 1$. ■

Proof of Theorem 5.1: Using Proposition 5.7 and Theorem 5.4 we have

$$\begin{aligned} D(\tilde{m}_n^-) &\leq C \left[C_1 \frac{\log n}{n^{1/2}} + C_2 \left(\frac{1}{(n)^{p/2}} + \frac{1}{(n\sigma^2)^p} \right)^{1/(2p+1)} \right] \\ &\leq C_3 \left(\frac{1}{n} \right)^{(1/4)(1-\delta(p))} \end{aligned} \quad (5.4)$$

where $\delta(p) = 1/(2p+1)$ and $C, C_1, C_2, C_3 > 0$. Since $\tilde{m}_n^- = {}_d m_n$, we have $\tilde{\mu}(\tilde{m}_n^-/\sqrt{n} \leq \sigma t) = \mu(m_n/\sqrt{n} \leq \sigma t)$ and consequently $D(\tilde{m}_n^-) = D(m_n)$ under the implied measures $\pi_* \tilde{\mu} = \mu$. It remains to deal with the coboundary. Observe

that for constants c and $\epsilon > 0$

$$\begin{aligned}
& \mu(v_n/\sqrt{n} \leq \sigma c) - \Phi(c) \\
& \leq \mu\left(\frac{1}{\sigma\sqrt{n}}m_n < c + \epsilon\right) + \mu\left(\frac{1}{\sigma\sqrt{n}}\underbrace{|v_n - m_n|}_{=\chi \circ T^n - \chi} \geq \epsilon\right) - \Phi(c) \\
& = \left[\mu\left(\frac{1}{\sigma\sqrt{n}}m_n < c + \epsilon\right) - \Phi(c + \epsilon)\right] + [\Phi(c + \epsilon) - \Phi(c)] , \tag{5.5}
\end{aligned}$$

where the third line follows by taking for example $\epsilon = (2\|\chi\|_\infty + 1)/\sigma\sqrt{n}$. The term in the first set of square brackets in (5.5) is the one that we have estimated in (5.4) whilst the second set of brackets gives $\mathcal{O}(\epsilon) = \mathcal{O}(n^{-1/2})$. Thus we deduce that there exists a constant $C > 0$ such that

$$\mu(v_n/\sqrt{n} \leq \sigma c) - \Phi(c) \leq Cn^{-(1/4)(1-\delta(p))} . \tag{5.6}$$

Similarly for the other direction we estimate that

$$\begin{aligned}
& \mu\left(\frac{1}{\sigma\sqrt{n}}v_n < c\right) - \Phi(c) \\
& \geq \mu\left(\frac{1}{\sigma\sqrt{n}}m_n < c - \epsilon\right) - \mu\left(\frac{1}{\sigma\sqrt{n}}|v_n - m_n| \geq \epsilon\right) - \Phi(c) \\
& = \left[\mu\left(\frac{1}{\sigma\sqrt{n}}m_n < c - \epsilon\right) - \Phi(c - \epsilon)\right] + [\Phi(c - \epsilon) - \Phi(c)] .
\end{aligned}$$

We therefore have that there exists a constant $C > 0$ such that

$$\mu(v_n/\sqrt{n} \leq \sigma c) - \Phi(c) \geq -Cn^{-(1/4)(1-\delta(p))} . \tag{5.7}$$

Hence by (5.6) and (5.7) we conclude that for all $\delta > 0$ there exists a constant $C > 0$ such that $D(v_n) \leq C n^{-1/4+\delta}$. ■

Chapter 6

Weak Invariance Principle

6.1 Weak Convergence and Donsker's Invariance Principle

Let $X, X_n, n \geq 1$ be random functions with values in the metric space (\mathcal{X}, d) where \mathcal{X} is a function space. Let $\mathbb{P}_X, \mathbb{P}_{X_n}$ denote the probability measures (laws) on the underlying space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ associated with X, X_n . That is, the probability law of X is the image probability measure $\mathbb{P} \circ X^{-1}$ induced by X on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, i.e. $\mathbb{P}_X(A) = \mathbb{P} \circ X^{-1}(A) = \mathbb{P}(X \in A)$ for $A \in \mathcal{B}(\mathcal{X})$. Recall, [9, 88], that we say that a sequence of probability measures \mathbb{P}_n *converge weakly* to a law \mathbb{P} , and write $\mathbb{P}_n \rightarrow_w \mathbb{P}$, if

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}} f d\mathbb{P}_n = \int_{\mathcal{X}} f d\mathbb{P} \quad \text{for all } f \in \mathcal{C}_b(\mathcal{X})$$

where $\mathcal{C}_b(\mathcal{X})$ is the set of all bounded, continuous, real-valued functions on S .

The random functions X_n *converge weakly* to X , $X_n \rightarrow_w X$, if the distributions \mathbb{P}_{X_n} converge weakly to a law \mathbb{P}_X i.e.

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}} f d\mathbb{P}_{X_n} = \int_{\mathcal{X}} f d\mathbb{P}_X \quad \text{for all } f \in \mathcal{C}_b(\mathcal{X}) \quad (6.1)$$

Equivalently, X_n converge weakly to X if

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad \text{for all } f \in \mathcal{C}_b(\mathcal{X}). \quad (6.2)$$

That is, convergence in distribution of random functions is the same as weak convergence of (the underlying) probability measures.

An important consequence of weak convergence, as it can be seen from the definitions (6.1)-(6.2), is that it is preserved under continuous mappings, that is if $X_n \rightarrow_w X$ then we have that $f(X_n) \rightarrow_w f(X)$ for any \mathcal{X} -measurable function f which is continuous almost everywhere with respect to \mathbb{P}_X , hence allowing us to obtain limit theorems for many related problems.

Remark 6.1. *The classical theory of weak convergence was developed primarily in the 1950s, with Prokhorov's fundamental paper [67] appearing in 1956. The standard exposition to weak convergence on $\mathcal{C}([0, 1])$ is Billingsley [9] and an extensive theory appears also in Parthasarathy [61]. Earlier developments can be found in Skorokhod [77], and Varadarajan [84]. An exhaustive treatment of weak convergence focusing on semimartingales is [42] whilst [65, 83] present an account on the topic of empirical processes.*

Let us now recall Donsker's Weak Invariance Principle: Assume that $\{X_i\}_{i \geq 1}$ is a sequence of independent and identically distributed random variables taking values on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[X_i] = 0$ and $\text{Var}[X_i] = 1$ for each $i \geq 1$. Let $S_0 = 0$ and consider the random walk $S_n = \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. Define

$$W_n(t) = \frac{1}{\sqrt{n}} [S_{[nt]} + (nt - [nt])X_{[nt]+1}]$$

for $t \in [0, 1]$ and $n \in \mathbb{N}$. In 1951, Donsker [20] proved that W_n converges in distribution to the standard Brownian motion W on $\mathcal{C}([0, 1])$ with respect to the sup-norm topology. Donsker's Theorem is the original version of an invariance principle where we have that the resulting limiting distribution is invariant of the specific distribution of the random variables X_i (as long as they are i.i.d. with mean zero and variance one). It is also known as the functional Central

Limit Theorem because it implies weak convergence for many functionals of interest.

In this chapter we prove the Weak Invariance Principle for ergodic stationary processes under the General Framework **(H)** having therefore, by Theorem 3.9, a martingale approximation at our disposal.

The continuous mapping theorem [54, 20, 24, 25, 36] is indispensable.

Proposition 6.2. (Continuous Mapping Theorem) *Let \mathcal{X}, \mathcal{Y} be metric spaces and consider random variables $\xi, \{\xi_n\}_{n \geq 1}$ with values in \mathcal{X} . Assume that $h : \mathcal{X} \rightarrow \mathcal{Y}$ is continuous. If $\xi_n \rightarrow_w \xi$ then $h(\xi_n) \rightarrow_w h(\xi)$.*

In particular, taking $\mathcal{X} = \mathcal{C}([0, 1])$, $\mathcal{Y} = \mathbb{R}$ and applying $h(f) = f(1)$ to $W_n \rightarrow_w W$, where W is standard Brownian motion, gives that $W_n(1) \rightarrow_w W(1) =_d \mathcal{N}(0, 1)$ hence retrieving the Central Limit Theorem.

6.2 The Weak Invariance Principle for Martingale Approximations

Throughout the section we denote by W a centered Brownian motion with variance $\sigma^2 = \int m^2 d\mu$ unless stated otherwise, where m is the function associated to v by Theorem 3.9. Define the continuous process $W_n \in \mathcal{C}([0, 1])$ by

$$W_n(t) = \frac{1}{\sqrt{n}} \left[\sum_{j=0}^{[nt]-1} v \circ T^j + (nt - [nt])v \circ T^{[nt]} \right]. \quad (6.3)$$

We prove the following:

Theorem 6.3. (Weak Invariance Principle) *Assume the T satisfies the General Framework **(H)**. Let $v : \Lambda \rightarrow \mathbb{R}$ be Hölder and of mean zero. Define W_n as in (6.3). Then $W_n \rightarrow_w W$ in $\mathcal{C}([0, 1])$.*

Remark 6.4. *Under the conditions used in this thesis, the Weak Invariance Principle was first proved by Hofbauer and Keller [38] in 1982 using different techniques. In fact a stronger statistical property (known as the almost sure invariance principle) is proved there for large classes of dynamical systems. For recent developments in this direction we turn to [51] and the references therein.*

Remark 6.5. *The theory for extending weak convergence for sequences of (continuous) random functions on $[0, 1]$ to the domain $[0, \infty)$ can be found in Whitt [87] which elaborates on earlier work done in Stone [78]. It follows from there that the proof of Theorem 6.3 works on any compact interval and therefore by establishing weak convergence in the interval $[0, S]$ for each fixed $S > 0$ one can extend the invariance principle in Theorem 6.3 to the real-halfline.*

The Weak Invariance Principle for stationary, ergodic martingale difference sequences was established by Billingsley [6] (see also Theorem 18.3 in [9]).

Theorem 6.6. (WIP, Billingsley) *Let $\{\xi_k\}_k$ be a stationary, ergodic martingale difference sequence with respect to some filtration $\{\mathcal{G}_k\}_k$ of \mathcal{F} on some probability space $(\Lambda, \mathcal{F}, \mu)$. Assume that $\mathbb{E}[\xi_n^2] = \sigma^2$ is positive and finite and let W be a standard Brownian motion. If we take*

$$X_n(t) = (\sigma\sqrt{n})^{-1} \sum_{k \leq [nt]-1} \xi_k + (nt - [nt])\xi_{[nt]}$$

then $X_n \rightarrow_w W$.

We begin by considering the continuous-time process corresponding to m_n . Define M_n on (Λ, μ) by

$$M_n(t) = n^{-1/2} \sum_{j=0}^{[nt]-1} m \circ T^j, \quad t \in \{i/n : 0 \leq i \leq n\}$$

and linearly interpolate to obtain

$$M_n(t) = n^{-1/2} \left(\sum_{j=0}^{[nt]-1} m \circ T^j + (nt - [nt])m \circ T^{[nt]} \right) \quad (6.4)$$

so that $M_n \in \mathcal{C}([0, 1])$.

We have seen in the previous chapter that when $v_n - m_n$ is sufficiently small then the Central Limit Theorem carries over from m_n to v_n . The corresponding result for the Weak Invariance Principle is given next.

Proposition 6.7. *Let W_n and M_n be given by (6.3) and (6.4) respectively. We have $\sup_t |W_n - M_n| \rightarrow 0$ a.e.*

Proof: By the triangle inequality we have

$$\begin{aligned} \sup_t |W_n(t) - n^{-1/2}v_{[nt]}| &\leq n^{-1/2} \max_{j \leq n} |v \circ T^j| \leq n^{-1/2} \|v\|_\infty, \\ \sup_t |M_n(t) - n^{-1/2}m_{[nt]}| &\leq n^{-1/2} \max_{j \leq n} |m \circ T^j| \leq n^{-1/2} \|m\|_\infty, \\ \sup_t |n^{-1/2}v_{[nt]} - n^{-1/2}m_{[nt]}| &= \sup_t |n^{-1/2}(\chi \circ T^{nt} - \chi)| \leq 2n^{-1/2} \|\chi\|_\infty. \end{aligned}$$

It follows that $\sup_t |W_n(t) - M_n(t)| \rightarrow 0$ a.e. ■

Hence by Proposition 6.7 we have that in proving Theorem 6.3 it suffices to prove the Weak Invariance Principle for M_n .

Theorem 6.8. (WIP for reverse m.d.s.'s) *The Weak Invariance Principle holds for M_n i.e.*

$$M_n \rightarrow_w W \quad \text{in } \mathcal{C}([0, 1]) \quad \text{as } n \rightarrow \infty.$$

Remark 6.9. *Since $\{m \circ T^j\}_{j \geq 0}$ is a sequence of reverse martingale differences we can not apply Theorem 6.6, the Weak Invariance Principle for martingales with stationary and ergodic differences, directly. We deal with this issue by passing, as we did in section 4.4 in proving the Central Limit Theorem, to the natural extension $\tilde{T} : \tilde{\Lambda} \rightarrow \tilde{\Lambda}$. This is an invertible map with ergodic*

invariant measure $\tilde{\mu}$, and there is a measurable projection $\pi : \tilde{\Lambda} \rightarrow \Lambda$ such that $\pi \circ \tilde{T} = T \circ \pi$ and $\pi_* \tilde{\mu} = \mu$. We obtain the lifted observable $\tilde{m} = m \circ \pi : \tilde{\Lambda} \rightarrow \mathbb{R}$ and we have that the joint distributions of $\{m \circ T^j : j \geq 0\}$ are identical to those of $\{\tilde{m} \circ \tilde{T}^j : j \geq 0\}$.

Anticipating the proof of Theorem 6.8 we introduce the continuous-time processes corresponding to \tilde{m}_n and \tilde{m}_n^- . Define \tilde{M}_n on $(\tilde{\Lambda}, \tilde{\mu})$ by

$$\tilde{M}_n(t) = n^{-1/2} \sum_{j=0}^{[nt]-1} \tilde{m} \circ \tilde{T}^j, \quad t \in \{i/n : 0 \leq i \leq n\}$$

and linearly interpolate to obtain

$$\tilde{M}_n(t) = n^{-1/2} \left(\sum_{j=0}^{[nt]-1} \tilde{m} \circ \tilde{T}^j + (nt - [nt]) \tilde{m} \circ \tilde{T}^{[nt]} \right)$$

so that $\tilde{M}_n \in \mathcal{C}([0, 1])$. Next define the backwards process

$$\tilde{M}_n^-(t) = n^{-1/2} \sum_{j=-[nt]}^{-1} \tilde{m} \circ \tilde{T}^j, \quad t \in \{i/n : 0 \leq i \leq n\}$$

and linearly interpolate to obtain

$$\tilde{M}_n^-(t) = n^{-1/2} \left(\sum_{j=-[nt]}^{-1} \tilde{m} \circ \tilde{T}^j + (nt - [nt]) \tilde{m} \circ \tilde{T}^{-[nt]-1} \right) \quad (6.5)$$

so that $\tilde{M}_n^- \in \mathcal{C}([0, 1])$.

Proposition 6.10. *We have*

$$\tilde{M}_n^- \rightarrow_w W \text{ in } \mathcal{C}([0, S]).$$

Proof: By Proposition 4.23 we have that (6.5) defines an ergodic, stationary martingale for which $\sigma^2 = \mathbb{E}[\tilde{m}^2] < \infty$. Since $[nt]/n \rightarrow t$ it then follows from Theorem 6.6 that $\widetilde{M}_n^- \rightarrow_w W$. ■

We proceed by relating weak convergence of \widetilde{M}_n^- and \widetilde{M}_n . First we note the following:

Lemma 6.11. *Consider the continuous functional $g : \mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ given by $g(u)(t) = u(1) - u(1 - t)$. Then $\text{Lip}(g) \leq 2$.*

Proof: Let $u, v \in \mathcal{C}([0, 1])$. Then

$$\begin{aligned} \sup_t |g(u)(t) - g(v)(t)| &= \sup_t |u(1) - u(1 - t) - v(1) + v(1 - t)| \\ &\leq \sup_t |u(1) - v(1)| + \sup_t |u(1 - t) - v(1 - t)| \\ &\leq 2 \sup_t |u(t) - v(t)| \end{aligned}$$

and therefore $\text{Lip}(g) \leq 2$. ■

Proposition 6.12. *We have*

$$\widetilde{M}_n \circ \widetilde{T}^{-n} = g(\widetilde{M}_n^-). \quad (6.6)$$

Proof: First we compute that

$$\widetilde{M}_n \circ \widetilde{T}^{-n}(t) = n^{-1/2} \left(\sum_{j=-n}^{[nt]-1-n} \tilde{m} \circ \widetilde{T}^j + (nt - [nt])\tilde{m} \circ \widetilde{T}^{[nt]-n} \right).$$

Moreover,

$$\widetilde{M}_n^-(1) = n^{-1/2} \sum_{j=-n}^{-1} \tilde{m} \circ \widetilde{T}^j$$

and

$$\widetilde{M}_n^-(1-t) = \frac{1}{\sqrt{n}} \left(\sum_{j=-[n(1-t)]}^{-1} \tilde{m} \circ \widetilde{T}^j + (n(1-t) - [n(1-t)])\tilde{m} \circ \widetilde{T}^{-[n(1-t)]-1} \right).$$

Let then $t > 0$ (for otherwise we, trivially, have that $g(\widetilde{M}_n^-)$ vanishes). Consider the following two cases:

Case 1: $t \neq \frac{\beta}{n}$ for an integer $\beta > 0$. Using $[n(1-t)] = n - [nt] - 1$ for $t > 0$ we compute that

$$\begin{aligned} g(\widetilde{M}_n^-)(t) &= \widetilde{M}_n^-(1) - \widetilde{M}_n^-(1-t) \\ &= n^{-1/2} \sum_{j=-n}^{-n+[nt]} \widetilde{m} \circ \widetilde{T}^j - n^{-1/2}([nt] - nt + 1) \widetilde{m} \circ \widetilde{T}^{-n+[nt]}. \end{aligned}$$

For ease of notation let us set $k = [nt] - n$. We have

$$\begin{aligned} &\widetilde{M}_n \circ \widetilde{T}^{-n} - g(\widetilde{M}_n^-) \\ &= \left[-\widetilde{m} \circ \widetilde{T}^k + (nt - [nt]) \widetilde{m} \circ \widetilde{T}^k - (nt - [nt] - 1) \widetilde{m} \circ \widetilde{T}^k \right] \frac{1}{\sqrt{n}} \\ &= 0. \end{aligned}$$

Case 2: Take $t = \frac{\beta}{n}$ where β is a positive integer. Then $[n(1-t)] = n - \beta$ and we have that

$$\widetilde{M}_n \circ \widetilde{T}^{-n}(t) = \frac{1}{\sqrt{n}} \sum_{j=-n}^{\beta-1-n} \widetilde{m} \circ \widetilde{T}^j.$$

Moreover,

$$\widetilde{M}_n^-(1-t) = \frac{1}{\sqrt{n}} \sum_{j=-n+\beta}^{-1} \widetilde{m} \circ \widetilde{T}^j.$$

Hence we deduce that

$$\begin{aligned} g(\widetilde{M}_n^-)(t) &= \widetilde{M}_n^-(1) - \widetilde{M}_n^-(1-t) \\ &= n^{-1/2} \sum_{j=-n}^{-1} \widetilde{m} \circ \widetilde{T}^j - \frac{1}{\sqrt{n}} \sum_{j=-n+\beta}^{-1} \widetilde{m} \circ \widetilde{T}^j \\ &= n^{-1/2} \sum_{j=-n}^{\beta-1-n} \widetilde{m} \circ \widetilde{T}^j \\ &= \widetilde{M}_n \circ \widetilde{T}^{-n}(t). \end{aligned}$$

■

Corollary 6.13. *We have*

$$\widetilde{M}_n \rightarrow_w g(W) \text{ in } \mathcal{C}([0, 1]) \text{ as } n \rightarrow \infty.$$

Proof: By Proposition 6.10 we have $\widetilde{M}_n^- \rightarrow_w W$. Hence by the continuous mapping theorem (Proposition 6.2) we have $g(\widetilde{M}_n^-) \rightarrow_w g(W)$. But Proposition 6.12 gives $\widetilde{M}_n =_d \widetilde{M}_n \circ \widetilde{T}^{-n} = g(\widetilde{M}_n^-)$. The result follows. ■

Lemma 6.14. *We have*

$$g(W) =_d W \text{ in } \mathcal{C}([0, 1]).$$

Proof: Notice that $g(W)(0) = W(1) - W(1) = 0$. Moreover, since W is sample path continuous we have that $g(W)$ is sample continuous as well. Take $0 \leq s < t \leq 1$. Then it follows by stationarity of Brownian increments that $g(W)(t) - g(W)(s) = W(1) - W(1-t) - (W(1) - W(1-s)) = W(1-s) - W(1-t) \sim \mathcal{N}(0, t-s)$. It remains to show that $g(W)$ has independent increments. Let $0 = t_1 < t_2 < \dots < t_n \leq 1$ and $n \geq 0$. Let $i = 1, \dots, n$ and set $1 - t_i = \hat{t}_{n+1-i}$. Then $\hat{t}_1 < \hat{t}_2 < \dots < \hat{t}_n$ and for all $1 \leq k \leq n-1$ we have

$$g(W)(t_{k+1}) - g(W)(t_k) = W(1 - t_k) - W(1 - t_{k+1}) = W(\hat{t}_{n-k+1}) - W(\hat{t}_{n-k}).$$

The assertion then follows from the independence of increments for the Brownian motion. ■

Proof of Theorem 6.8: Since $\widetilde{M}_n = M_n \circ \pi =_d M_n$ it suffices to prove that $\widetilde{M}_n \rightarrow_w W$. But Corollary 6.13 and Lemma 6.14 give that

$$M_n =_d \widetilde{M}_n \rightarrow_w g(W) =_d W$$

in $\mathcal{C}([0, 1])$. ■

Proof of Theorem 6.3: Note that $\sup_{t \in [0,1]} |W_n(t) - M_n(t)| \rightarrow 0$ almost surely by Proposition 6.7. The result follows then from Theorem 6.8. ■

Chapter 7

Rate of convergence in the Weak Invariance Principle

7.1 Introduction and Statement of the Theorem

The *Lévy-Prokhorov distance*, [67], $\pi_1(X, Y)$ between two processes $X, Y \in \mathcal{C}$, is defined by

$$\pi_1(X, Y) = \inf \{ \epsilon > 0 : \mathbb{P}_X(A) \leq \mathbb{P}_Y(A^\epsilon) + \epsilon \text{ for all closed } A \in \mathcal{B}(\mathcal{C}) \} . \quad (7.1)$$

Here we write $\mathcal{C} = \mathcal{C}([0, 1])$ and $\mathcal{B}(\mathcal{C})$ is the Borel σ -algebra of \mathcal{C} . The set $A^\epsilon = \{ \omega : \rho(A, \omega) \leq \epsilon \}$ is the ϵ -neighbourhood of A where

$$\rho(A, \omega) = \inf_{\omega' \in A} \rho(\omega', \omega)$$

and $\rho(\cdot, \cdot)$ denotes the uniform distance,

$$\rho(x, y) = \sup_{t \in [0, 1]} |x(t) - y(t)| , \quad x, y \in \mathcal{C}([0, 1]) .$$

It is known, Borovkov [12], that for the i.i.d. r.v.'s case (Donsker's Invariance Principle, section 6.1) we have that $\pi_1(W_n, W)$ is $\mathcal{O}(n^{-1/4})$ and this rate is optimal [2].

Remark 7.1. *The Lévy-Prokhorov distance was introduced by Prokhorov in his seminal paper [67]. It is an extension of the Lévy metric for measures on the real line. Though not immediately obvious, π_1 is symmetric (e.g. Dudley [22] who writes that this was known to Strassen [79] – a proof can also be found in [24, 23, 28]). It is standard that π_1 is a metric and induces the topology of weak convergence on the set of all probability measures on a metric space \mathcal{X} as proved in [67] for \mathcal{X} complete and for general separable \mathcal{X} in [22]). It is obvious from the definition (7.1) that the Lévy-Prokhorov metric computes distances of probability measures (laws). For the sake of simplicity we use the notation $\pi_1(X, Y)$ which should be interpreted as meaning $\pi_1(\mathbb{P}_X, \mathbb{P}_Y)$.*

Remark 7.2. *There are other metrics inducing the weak topology: In terms of the space of Lipschitz functions, section 3.2, consider Borel probability measures \mathbb{P} and \mathbb{Q} on a metric space (S, d) and define*

$$\beta(\mathbb{P}, \mathbb{Q}) = \sup \left\{ \left| \int f d(\mathbb{P} - \mathbb{Q}) \right| : \|f\|_{\text{Lip}} \leq 1 \right\}.$$

The bounded Lipschitz metric, β , was defined by Dudley in [21] and he has shown that for any metric space (S, d) , β is a metric on the set of all laws on S , [24]. A collection of other metrics that metrize weak convergence (under a variety of assumptions) can be found in [30] (Theorem 6 in particular). Lastly, recall the Ky-Fan metric α for random variables X and Y , e.g. [92], defined by

$$\alpha(X, Y) = \inf \{ \epsilon > 0 : \mathbb{P}[d(X, Y) > \epsilon] \leq \epsilon \}.$$

This is the natural topology of convergence in probability. Then, [24] (Theorem 11.3.5), for any separable metric space (S, d) and random variables X and Y on S we have $\pi_1(X, Y) \leq \alpha(X, Y)$, hence providing a probabilistic interpretation of the Lévy-Prokhorov metric (and also showing that convergence in probability implies convergence in distribution). A related (converse) more

refined result, useful in deducing estimates in the Lévy-Prokhorov metric, is the Strassen-Dudley Theorem [24] (in particular Corollary 11.6.4 therein).

Now, let us recall that the linearly interpolated process (6.3) gives a continuous random polygonal line such that each trajectory lies in $\mathcal{C}([0, 1])$. Since the topology of weak convergence of probability laws on the Borel σ -algebra of $\mathcal{C}([0, 1])$ is metrizable by the Lévy-Prokhorov metric [24] and $X_n \rightarrow_w X$ as $n \rightarrow \infty$ is equivalent to $\pi_1(X_n, X) \rightarrow 0$ as $n \rightarrow \infty$ we can estimate the rate of convergence in the WIP (Theorem 6.3) by estimating the rate of convergence in this metric between W_n and a Brownian motion W . In particular we prove the following:

Theorem 7.3. (Rate of Convergence in the WIP) *Assume that T satisfies the General Framework (H) and $v \in \mathcal{C}^\alpha(\Lambda)$ with $\int_\Lambda v d\mu = 0$. Let W_n be defined by (6.3) and W be a centered Brownian motion with variance $\sigma^2 = \int m^2 d\mu$, where m is the function associated to v by Theorem 3.9. Then for all $\delta > 0$ there exists a constant $C > 0$ such that*

$$\pi_1(W_n, W) \leq Cn^{-1/4+\delta}.$$

7.2 Error Rate in the WIP for martingale approximations

Let us first note a property of the Lévy-Prokhorov distance.

Proposition 7.4. *Let $M_n, L_n \in \mathcal{C}([0, 1])$, $q > 1$ Assume that*

$$\left\| \sup_{t \in [0, 1]} |M_n(t) - L_n(t)| \right\|_q \leq C_1 n^{-p}$$

for some constants $C_1 > 0$ and $p > 0$. Then there exists a constant $C_2 > 0$ such that for all $n \geq 1$ we have $\pi_1(M_n, L_n) \leq C_2 n^{-pq/(q+1)}$.

Proof: Observe that by Markov's inequality we have

$$\mathbb{P} \left[\sup_{t \in [0,1]} |M_n(t) - L_n(t)| > \epsilon \right] \leq \epsilon^{-q} \left\| \sup_{t \in [0,1]} |M_n(t) - L_n(t)| \right\|_q^q \leq C_1^q n^{-pq} \epsilon^{-q}.$$

Choosing $\epsilon = C_1^{q/(q+1)} n^{-pq/(q+1)}$ we deduce that

$$\inf \left\{ \delta > 0 : \mathbb{P} \left[\sup_{t \in [0,1]} |M_n(t) - L_n(t)| > \delta \right] \leq \delta \right\} \leq \epsilon.$$

Let $A \in \mathcal{B}(\mathcal{C})$. If $M_n \in A$ and $\sup_{t \in [0,1]} |M_n(t) - L_n(t)| < \epsilon$ then we have by the definition of A^ϵ that $L_n \in A^\epsilon$. Hence

$$\mathbb{P}_{M_n}(A) = \mathbb{P}[M_n \in A] \leq \mathbb{P}[L_n \in A^\epsilon] + \epsilon = \mathbb{P}_{L_n}(A^\epsilon) + \epsilon$$

and we conclude that $\pi_1(M_n, L_n) \leq \epsilon$. ■

Let V_n denote the conditional variance as in (5.1). Define

$$X_n(t) = n^{-1/2} \sum_{j=1}^k \tilde{m} \circ \tilde{T}^{-j}, \quad \text{for } t = (V_n)^{-1} V_i, \quad i = 0, 1, \dots, n.$$

Linearly interpolating on the subintervals $[(V_k)^{-1} V_{j-1}, (V_k)^{-1} V_j]$, $V_0 = 0$, $j = 1, \dots, k$ we denote the partial sums process X_n by

$$X_n(t) = n^{-1/2} \left(\tilde{m}_k^- + (V_{k+1} - V_k)^{-1} (tV_n - V_k) \tilde{m} \circ \tilde{T}^{-k-1} \right) \\ \text{if } V_k \leq tV_n < V_{k+1} \quad (7.2)$$

for $t \in [0, 1]$, $0 \leq k \leq n$ i.e. the random element X_n with values in $\mathcal{C}([0, 1])$, the space of real-valued continuous functions on the unit interval endowed with the topology of uniform convergence, $\rho(\cdot, \cdot)$, and the corresponding Borel σ -algebra.

Bounds on the Lévy-Prokhorov distance between the continuous process X_n constructed from square-integrable martingale differences as in (7.2) and a Brownian motion W are given in Kublicius [52] (Corollary). Note that since

$\|m\|_\infty < \infty$ the hypothesis in the Corollary is satisfied and in our setting the result is reduced to Lemma 7.5 below.

Lemma 7.5. *Consider $X_n \in \mathcal{C}$ as in (7.2) and W a Brownian motion with variance σ^2 . There exists a real constant $C > 0$ such that*

$$\pi_1(X_n, W) \leq C \left\{ n^{-1/4} + \inf_{0 \leq \epsilon \leq 1} \left\{ \epsilon + \mathbb{P}[|(n\sigma^2)^{-1}V_n - 1| > \epsilon^2] \right\} \right\} \ln n.$$

Proposition 7.6. *For all $\delta > 0$ there exists a constant $C > 1$ such that*

$$\pi_1(X_n, W) \leq C n^{-1/4+\delta}.$$

Proof: Let $p \geq 2$. Applying Markov's inequality and using Proposition 5.7 we obtain

$$\mathbb{P} \left[\left| \frac{V_n}{n\sigma^2} - 1 \right| > \epsilon^2 \right] \leq \epsilon^{-2p} \mathbb{E} \left| \frac{V_n}{n\sigma^2} - 1 \right|^p = \epsilon^{-2p} \left\| \frac{V_n}{n\sigma^2} - 1 \right\|_p^p \leq C \epsilon^{-2p} n^{-p/2}.$$

Choosing $\epsilon = n^{-p/(4p+2)}$ in Lemma 7.5 and setting $\tilde{\delta} := \tilde{\delta}(p) = 1/4(2p+1)$ gives

$$\pi_1(X_n, W) \leq C n^{-1/4+\tilde{\delta}} \ln n$$

where $C > 0$. Note that $\ln x < x$ for all $x > 0$ and therefore for all $\beta > 0$ we have $\ln x = \frac{1}{\beta} \ln x^\beta < \frac{1}{\beta} x^\beta$. Letting $\delta = 2\tilde{\delta}$ we deduce that

$$\pi_1(X_n, W) \leq C n^{-1/4+\delta}.$$

■

Recall that under the General Framework **(H)** we have deduced in Proposition 4.23 that \tilde{m}_n^- martingale with respect to $\{\tilde{T}^j \tilde{\mathcal{F}}_j\}$. We now relate the process \tilde{M}_n^- defined in (6.5) to X_n , noting that (7.2) introduces an integer-valued random variable k by means of an averaging condition. We use this to find a quantitative connection between the two processes.

Lemma 7.7. Define $K_n = \max_{i \leq n^{1/2}} \max_{\ell \leq n^{1/2}} \left| \sum_{j=in^{1/2}}^{in^{1/2}+\ell} \tilde{m} \circ \tilde{T}^{-j} \right|$, $n \geq 1$. Then

$$(a) \left| \sum_{j=a}^{b-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq K_n (n^{-1/2}(b-a) + 3) \text{ for all } 0 \leq a < b \leq n.$$

$$(b) \|K_n\|_p \leq C \|m\|_p n^{\frac{1}{4} + \frac{1}{2p}} \text{ for all } p \geq 2.$$

Proof: (a) Assume first that $b-a \geq \sqrt{n}$. Let i be the least integer such that $[i\sqrt{n}] > a$ and r be the biggest integer such that $[r\sqrt{n}] < b$. Let us split the sum into parts:

$$\left| \sum_{j=a}^{b-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq \left| \sum_{j=a}^{[i\sqrt{n}]-1} \tilde{m} \circ \tilde{T}^{-j} \right| + \left| \sum_{j=[i\sqrt{n}]}^{[r\sqrt{n}]-1} \tilde{m} \circ \tilde{T}^{-j} \right| + \left| \sum_{j=[r\sqrt{n}]}^{b-1} \tilde{m} \circ \tilde{T}^{-j} \right|. \quad (7.3)$$

Consider the second term in the RHS of (7.3). We have

$$\left| \sum_{j=[i\sqrt{n}]}^{[r\sqrt{n}]-1} \tilde{m} \circ \tilde{T}^{-j} \right| = \left| \sum_{m=i}^{r-1} \sum_{j=[m\sqrt{n}]}^{[(m+1)\sqrt{n}]-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq \sum_{m=i}^{r-1} \left| \sum_{j=[m\sqrt{n}]}^{[(m+1)\sqrt{n}]-1} \tilde{m} \circ \tilde{T}^{-j} \right|. \quad (7.4)$$

It is trivial that $[(m+1)\sqrt{n}] - 1 - [m\sqrt{n}] \leq [\sqrt{n}]$. Thus we have that (7.4) is bounded by

$$\sum_{m=i}^{r-1} \max_{\ell \leq \sqrt{n}} \left| \sum_{j=[m\sqrt{n}]}^{[m\sqrt{n}]+\ell} \tilde{m} \circ \tilde{T}^{-j} \right| \leq (r-i)K_n \leq \frac{b-a}{\sqrt{n}} K_n \quad (7.5)$$

where the last inequality follows by our choice of i and r .

Next, since r is the biggest integer such that $b > [r\sqrt{n}]$ we have that $b-1-[r\sqrt{n}] \leq [\sqrt{n}]$. Hence

$$\left| \sum_{j=[r\sqrt{n}]}^{b-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq \max_{\ell \leq \sqrt{n}} \left| \sum_{j=[r\sqrt{n}]}^{[r\sqrt{n}]+\ell} \tilde{m} \circ \tilde{T}^{-j} \right| \leq K_n.$$

Lastly observe that

$$\left| \sum_{j=a}^{\lfloor i\sqrt{n} \rfloor - 1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq \left| \sum_{j=\lfloor (i-1)\sqrt{n} \rfloor}^{\lfloor i\sqrt{n} \rfloor - 1} \tilde{m} \circ \tilde{T}^{-j} \right| + \left| \sum_{j=\lfloor (i-1)\sqrt{n} \rfloor}^{a-1} \tilde{m} \circ \tilde{T}^{-j} \right|.$$

The first term in the LHS in the equation above was shown in (7.5) to be bounded by K_n . Since $a - 1 - \lfloor (i-1)\sqrt{n} \rfloor < \lfloor i\sqrt{n} \rfloor - \lfloor (i-1)\sqrt{n} \rfloor - 1 \leq \lfloor \sqrt{n} \rfloor$ we deduce that

$$\left| \sum_{j=\lfloor (i-1)\sqrt{n} \rfloor}^{a-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq \max_{\ell \leq \sqrt{n}} \left| \sum_{j=\lfloor (i-1)\sqrt{n} \rfloor}^{\lfloor (i-1)\sqrt{n} \rfloor + \ell} \tilde{m} \circ \tilde{T}^{-j} \right| \leq K_n$$

which confirms the claim for $b - a \geq \sqrt{n}$.

Let us now assume that $b - a < \sqrt{n}$ and let $a \leq \lfloor i\sqrt{n} \rfloor < b$. Then

$$\left| \sum_{j=a}^{b-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq \left| \sum_{j=a}^{\lfloor i\sqrt{n} \rfloor - 1} \tilde{m} \circ \tilde{T}^{-j} \right| + \left| \sum_{j=\lfloor i\sqrt{n} \rfloor}^{b-1} \tilde{m} \circ \tilde{T}^{-j} \right|$$

and by the previous argument we obtain the bound $3K_n$.

(b) Note that by measure invariance we have

$$\int |K_n|^p d\tilde{\mu} \leq \sum_{i \leq \sqrt{n}} \int \max_{\ell \leq \sqrt{n}} \left| \sum_{j=0}^{\ell} \tilde{m} \circ \tilde{T}^{-j-i\sqrt{n}} \right|^p d\tilde{\mu} \leq \sqrt{n} \left\| \max_{\ell \leq \sqrt{n}} \left| \sum_{j=0}^{\ell} \tilde{m} \circ \tilde{T}^{-j} \right| \right\|_p^p.$$

Taking p -th roots and applying Burkholder's inequality (Proposition 5.5) gives

$$\|K_n\|_p \leq C(\sqrt{n})^{1/p}(\sqrt{n})^{1/2} \|\tilde{m}\|_p$$

for some real constant $C > 0$. Since $\|\tilde{m}\|_p = \|m\|_p$ the result follows. \blacksquare

Proposition 7.8. *For all $\delta > 0$ there exists a constant C such that*

$$\pi_1(\tilde{M}_n^-, W) \leq Cn^{-1/4+\delta}.$$

Proof: Recall that for $t \in [0, 1]$ and $n \geq 1$ the random variable $k = k_{n,t}(x) \in \mathbb{N}$ is defined by the equation

$$V_k \leq tV_n < V_{k+1} \quad (7.6)$$

or equivalently by $k = \max \{i \in \mathbb{N} : V_i \leq tV_n\}$. Define the integer-valued random variable $\ell = \ell_{n,t}(x)$ by $\ell = k - [nt]$ and set $\tilde{V}_n = (\sigma^2)^{-1}V_n - n$. From equation (7.6) we obtain

$$\ell + \tilde{V}_k \leq t\tilde{V}_n + nt - [nt] < \ell + 1 + \tilde{V}_{k+1}.$$

As $0 \leq nt - [nt] \leq 1$ we have

$$\ell \leq t\tilde{V}_n - \tilde{V}_k + 1 \leq |t\tilde{V}_n - \tilde{V}_k + 1| \leq |t\tilde{V}_n| + |\tilde{V}_k| + 1 \leq |\tilde{V}_n| + |\tilde{V}_k| + 1$$

and

$$\ell \geq -(\tilde{V}_{k+1} + 1 - t\tilde{V}_n) \geq -|\tilde{V}_{k+1} + 1 - t\tilde{V}_n| \geq -(|\tilde{V}_n| + |\tilde{V}_{k+1}| + 1).$$

Since $k \leq n$ we have

$$|\ell| \leq 2 \max_{j \leq n+1} |\tilde{V}_j| + 1$$

and therefore

$$\sup_{t \in [0,1]} |k - [nt]| \leq 2 \max_{j \leq n+1} |\tilde{V}_j| + 1.$$

For $p \geq 2$ it follows from Proposition 5.7 that

$$\left\| \sup_{t \in [0,1]} |k - [nt]| \right\|_p \leq C_p(n+1)^{1/2} + 1 \leq Cn^{1/2} \quad (7.7)$$

for some constant $C > 0$. Now from Lemma 7.7 (a) we have that

$$\left| \sum_{j=k}^{[nt]-1} \tilde{m} \circ \tilde{T}^{-j} \right| \leq K_n(n^{-1/2}([nt] - k) + 3).$$

Raising to the power of $p/2$ gives

$$\left| \sum_{j=k}^{[nt]-1} \tilde{m} \circ \tilde{T}^{-j} \right|^{p/2} \leq |K_n|^{p/2} (n^{-1/2}([nt] - k) + 3)^{p/2}.$$

Taking expectations and applying Cauchy-Schwarz we obtain

$$\begin{aligned} \int \sup_{t \in [0,1]} \left| \sum_{j=k}^{[nt]-1} \tilde{m} \circ \tilde{T}^{-j} \right|^{p/2} d\tilde{\mu} &\leq \int |K_n|^{p/2} \sup_{t \in [0,1]} |(n^{-1/2}([nt] - k) + 3)^{p/2} d\tilde{\mu} \\ &\leq \left(\int |K_n|^p d\tilde{\mu} \right)^{\frac{p}{2p}} \left(\int \sup_{t \in [0,1]} |(n^{-1/2}([nt] - k) + 3)|^p d\tilde{\mu} \right)^{\frac{p}{2p}}. \end{aligned}$$

Raising to the power of $2/p$ gives

$$\left\| \sup_{t \in [0,1]} \left| \sum_{j=k}^{[nt]-1} \tilde{m} \circ T^{-j} \right| \right\|_{p/2} \leq \|K_n\|_p \left\| \sup_{t \in [0,1]} |n^{-1/2}([nt] - k) + 3| \right\|_p$$

and from (7.7) and Lemma 7.7 (b) we deduce that

$$\left\| \sup_{t \in [0,1]} \left| \sum_{j=k}^{[nt]-1} \tilde{m} \circ T^{-j} \right| \right\|_{p/2} \leq C n^{\frac{1}{4} + \frac{1}{2p}}.$$

Noting that

$$\frac{1}{\sqrt{n}} \sum_{j=k}^{[nt]-1} \tilde{m} \circ \tilde{T}^{-j} \circ \tilde{T}^{-1} = \frac{1}{\sqrt{n}} \sum_{j=k+1}^{[nt]} \tilde{m} \circ \tilde{T}^{-j} = \frac{1}{\sqrt{n}} \tilde{m}_{[nt]}^- - \frac{1}{\sqrt{n}} \tilde{m}_k^-$$

it then follows from Proposition 7.4 that $\pi_1(\tilde{M}_n^-, X_n) \leq C n^{-1/4+\delta}$. Thus we conclude by Proposition 7.6 that $\pi_1(\tilde{M}_n^-, W) \leq C n^{-1/4+\delta}$. ■

Let (Ω, d) be a metric space and recall that $h : (\Omega, d) \rightarrow (\Omega, d)$ is Lipschitz continuous if there exists a constant K such that

$$d(h(x), h(y)) \leq Kd(x, y) \quad \text{for all } x, y \in \Omega.$$

We need the following mapping theorem [86, 88].

Theorem 7.9. (Lipschitz Mapping Theorem) *Suppose that $h : (\Omega, \mu) \rightarrow (\Omega, \mu)$ is Lipschitz continuous with constant K on $A \subset \Omega$. Then*

$$\pi_1(h(X), h(Y)) \leq (K \vee 1)\pi_1(X, Y)$$

for any random elements X and Y of (Ω, μ) for which $\mathbb{P}[Y \in A] = 1$.

Proposition 7.10. *For all $\delta > 0$ there exists a constant $C > 1$ such that*

$$\pi_1(M_n, W) \leq Cn^{-1/4+\delta}.$$

Proof: We will apply Theorem 7.9 for the continuous functional g introduced in Lemma 6.11. Since $\widetilde{M}_n \circ \widetilde{T}^{-nS} = g(\widetilde{M}_n^-)$ (Proposition 6.12) and by the Lipschitz Mapping Theorem we have that $\pi_1(g(\widetilde{M}_n^-), g(W)) \leq C\pi_1(\widetilde{M}_n^-, W)$. Therefore we obtain

$$\begin{aligned} \pi_1(\widetilde{M}_n, W) &\leq \pi_1(\widetilde{M}_n, g(W)) = \pi_1(g(\widetilde{M}_n^-), g(W)) \\ &\leq C\pi_1(\widetilde{M}_n^-, W) \leq Cn^{-1/4+\delta} \end{aligned}$$

by Proposition 7.8. Since $\widetilde{M}_n = M_n \circ \pi$ we deduce that $\pi_1(M_n, W) = \pi_1(\widetilde{M}_n, W)$ and the result follows. ■

Proof of Theorem 7.3: By Proposition 7.10 we conclude that

$$\begin{aligned} \pi_1(W_n, W) &\leq \pi_1(W_n, M_n) + \pi_1(M_n, W) \\ &\leq C_1n^{-1/2} + C_2n^{-1/4+\delta} \leq Cn^{-1/4+\delta} \end{aligned}$$

for some positive real constants C_1, C_2, C . ■

Chapter 8

Fast-slow systems in Continuous Time

8.1 Introduction

In the final two chapter of this thesis we are investigating rates of convergence in homogenization for fast-slow systems, building on the work of [55, 32] which establishes homogenization results under the assumption that the slow dynamics is one-dimensional (in the additive and multiplicative noise cases) for very general fast dynamics. This will be made more precise in the next section.

First we offer a minimal, for the purpose of this thesis, discussion of stochastic integrals, the very rich theory of which can be widely found in many standard textbooks and monographs in the literature e.g. [41, 44, 45, 69, 70]. We follow closely [62, 63].

First recall that a *filtration* $\{\mathcal{F}_t\}_t$ on (Ω, \mathcal{F}) is a family of sub- σ -algebras of \mathcal{F} such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ where $t \geq s$. The filtration $\{\mathcal{F}_t^X\}_t$ generated by a stochastic process $X = \{X_t\}$ is $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$. A stochastic process $\{X_t\}_t$ is *adapted* to the filtration $\{\mathcal{F}_t\}_t$ if for all t we have that X_t is a \mathcal{F}_t -measurable random variable.

Assume now that $f(t)$ is a random process, adapted to the filtration $\{\mathcal{F}_t^W\}_{t \geq 0}$ generated by W where $W = \{W_t\}_{t \geq 0}$ is Brownian motion. Further assume that

$$\mathbb{E} \left(\int_0^T f(s)^2 ds \right) < \infty .$$

The *Itô stochastic integral* $I(t)$ is defined as the \mathcal{L}^2 limit of the following Riemann sum,

$$I(t) = \lim_{K \rightarrow \infty} \sum_{k=1}^{K-1} f(t_{k-1}) (W_{t_k} - W_{t_{k-1}}) ,$$

where $t_k = k\Delta t$ and $K\Delta t = t$ define a partition of the interval $[0, T]$. The resulting integral is written as

$$I(t) = \int_0^t f(s) dW_s .$$

It is standard that the Itô integral is a martingale,

$$\mathbb{E} [I(t)] = 0 ,$$

and

$$\mathbb{E} [I(t) \mid \mathcal{F}_s^W] = I(s) \text{ for all } t \geq s .$$

The most common alternative to the Itô integral is the *Stratonovich stochastic integral* $I_S(t)$ defined as the \mathcal{L}^2 limit of the Riemann sum,

$$I_S(t) = \lim_{K \rightarrow \infty} \sum_{k=1}^{K-1} \frac{1}{2} (f(t_k) + f(t_{k-1})) (W_{t_k} - W_{t_{k-1}}) ,$$

where $t_k = k\Delta t$ and $K\Delta t = t$. The resulting integral is denoted by

$$I_S(t) = \int_0^t f(s) \circ dW_s .$$

Remark 8.1. *In general, the Itô and the Stratonovich integral are different (though there are cases in which we can convert one stochastic integral into the other using a "Itô-to-Stratonovich correction" [62] and also cases that the*

two integrals coincide; we do not touch on these issues here). One difference is that the Stratonovich integral obeys the usual rules of ordinary calculus. In particular, the ordinary chain rule holds for the Stratonovich integral whilst for the Itô integral we have the more complex Itô's lemma [70].

A simple example illustrating the difference between the two integrals is when the integrand is a Brownian motion. Assume that W is a standard Brownian motion. The Stratonovich integral in this case is

$$\int_0^t W_s \circ dW_s = \frac{1}{2} W_t^2, \quad (8.1)$$

whilst the Itô integral requires a correction term,

$$\int_0^t W_s dW_s = \frac{1}{2} (W_t^2 - t). \quad (8.2)$$

It is a standard result that a Brownian motion is nowhere differentiable almost surely. If it was differentiable then we would have had that

$$\int_0^T W_s dW_s = \frac{1}{2} \int_0^T \frac{d}{ds} W_s^2 ds = \frac{1}{2} W_T^2.$$

This is the answer we get by the usual rules of ordinary calculus and this is what the Stratonovich integral gives in (8.1). Also W_t^2 is not a martingale since for $t > s$ we have

$$\mathbb{E} [W_t^2 \mid \mathcal{F}_s^W] = W_s^2 + (t - s).$$

But Itô integrals are martingales and hence have mean zero which is compatible with (8.2).

We now discuss briefly stochastic differential equations (SDEs). An Itô SDE is of the form

$$dX = A(X)dt + B(X)dW, \quad X(0) = x_0,$$

where A, B are some measurable functions with suitable regularity assumptions and $x_0 \in \mathbb{R}$. This is the notationally convenient differential form, the precise

meaning of which is that X_t satisfies the integral equation

$$X_t = x_0 + \int_0^t A(X_s) ds + \int_0^t B(X_s) dW_s, \quad t \geq 0.$$

A Stratonovich SDE is of the form

$$dX = A(X)dt + B(X) \circ dW, \quad X(0) = x_0.$$

In integral form this means that X_t satisfies

$$X_t = x_0 + \int_0^t A(X_s) ds + \int_0^t B(X_s) \circ dW_s, \quad t \geq 0.$$

Now, to motivate the problem we examine in chapters 8 and 9, consider a fast-slow system of ordinary differential equation (ODEs) of the form

$$\dot{x}_\epsilon = a(x_\epsilon, y_\epsilon) + \epsilon^{-1} b(x_\epsilon) v(y_\epsilon), \quad x_\epsilon(0) = \xi, \quad (8.3)$$

$$\dot{y}_\epsilon = \epsilon^{-2} g(y_\epsilon), \quad y_\epsilon(0) = \eta, \quad (8.4)$$

(ignoring many technicalities and conditions that are stated precisely in Assumptions 8.3 below). If b is the identity map we will speak of the *additive noise* case and otherwise we will refer to a *multiplicative noise* case. Assume that the fast dynamics in (8.4) is chaotic and induce white noise in the slow variables x_ϵ as a natural scale ϵ tends to zero. More suggestively, (we can) write the slow variables (8.3) in the form

$$dx_\epsilon = a(x_\epsilon, y_\epsilon)dt + b(x_\epsilon)dW_\epsilon, \quad x_\epsilon(0) = \xi, \quad (8.5)$$

where W_ϵ converges weakly to a Brownian motion W . Then the limiting dynamics of the slow variables converges weakly to a solution X to a SDE of the form

$$dX = a(X)dt + b(X) * dW, \quad X_\epsilon(0) = \xi. \quad (8.6)$$

Here, we have used the unusual notation to emphasize the issue on how to interpret the stochastic integral $\int B(X) * dW$ (we comment on this below). We refer to derivation of equation (8.6) as *homogenization*.

Let x_ϵ denote solutions of the ordinary differential equation

$$dx_\epsilon = a(x_\epsilon)dt + b(x_\epsilon)dW_\epsilon,$$

where W_ϵ converges weakly to Brownian motion W and a, b are smooth. The classical Wong-Zakai theorem [90, 91] gives sufficient conditions under which x_ϵ converges weakly to solutions X of (8.6) provided that the stochastic integral is given the Stratonovich interpretation, $\int B(X) * dW = \int B(X) \circ dW$. These conditions hold automatically in one dimension, which is the case we consider in this thesis (in higher dimensions the correct interpretation of the stochastic integral $\int B(X) * dW$ is more complex [81, 82]). Recent progress in smooth approximation of stochastic differential equations appears in [47].

The rest of the thesis is devoted to investigating the rates of convergence in homogenization of fast-slow systems.

8.2 Setting and Assumptions

Let a flow $\phi_t : \mathbb{R}^\ell \rightarrow \mathbb{R}^\ell$ be generated by an ODE $\dot{y} = g(y)$. We assume that there is a compact set $\Lambda \subset \mathbb{R}^\ell$ such that $\phi_t(y) \in \Lambda$ for all $t \geq 0$, for all $y \in \Lambda$, that is Λ is an invariant set for ϕ_t . Also, we suppose that there exists a probability measure μ on Λ such that $\mu(\phi_t E) = \mu(E)$ for all t and for all measurable sets E i.e. it is invariant. Moreover, let μ be an ergodic measure i.e. if $E \subset \Lambda$ is invariant then $\mu(E) \in \{0, 1\}$.

Introduce the observable $v : \Lambda \rightarrow \mathbb{R}$ and define $W_\epsilon \in \mathcal{C}([0, 1])$ by

$$W_\epsilon(t) = \epsilon v_{t\epsilon^{-2}}, \quad v_t = \int_0^t v \circ \phi_s ds.$$

We first consider the fast-slow system of ODEs with additive noise

$$\dot{x}_\epsilon = a(x_\epsilon, y_\epsilon) + \epsilon^{-1}v(y_\epsilon), \quad x_\epsilon(0) = \xi, \quad (8.7)$$

$$\dot{y}_\epsilon = \epsilon^{-2}g(y_\epsilon), \quad y_\epsilon(0) = \eta, \quad (8.8)$$

where $x_\epsilon \in \mathbb{R}$, $y_\epsilon \in \mathbb{R}^\ell$, under conditions that are stated precisely in Assumptions 8.3 below. Later on we will focus on the multiplicative noise case considering the fast-slow system (8.9), (8.8) where

$$\dot{x}_\epsilon = a(x_\epsilon, y_\epsilon) + \epsilon^{-1}b(x_\epsilon)v(y_\epsilon), \quad x_\epsilon(0) = \xi \quad (8.9)$$

and $b : \mathbb{R} \rightarrow \mathbb{R}$ satisfies some regularity conditions.

Write $\bar{a}(x) = \int_{\Lambda} a(x, y) d\mu(y)$ and let X denote the unique solution to the SDE

$$dX = \bar{a}(X)dt + dW, \quad X(0) = \xi. \quad (8.10)$$

It was shown in [55], Theorem 1.1, that for the additive noise case if $W_\epsilon \rightarrow_w W$ then $x_\epsilon \rightarrow_w X$ also. In the next section we obtain the corresponding rate of convergence. The multiplicative noise case will then be addressed in section 8.4, with the limiting SDE being of Stratonovich type as shown in [32] Theorem 3.3.

8.3 Rates of Convergence in the Additive Noise case for Flows

We prove the following:

Theorem 8.2. *Let Assumptions 8.3 below hold. Let (x_ϵ, y_ϵ) be the unique solution of the fast-slow system of ODEs (8.7), (8.8) and consider the SDE (8.10). Then we have $\pi_1(x_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$.*

We make the following standing assumptions.

Assumptions 8.3. *The fast-slow system (8.7), (8.8) satisfies the following:*

- *The observable $v : \Lambda \rightarrow \mathbb{R}$ is Hölder continuous with $\int v d\mu = 0$.*
- *$\xi \in \mathbb{R}$ is fixed throughout and $\eta \in \Lambda$ is the sole source of randomness in the fast-slow system.*
- *$a : \mathbb{R} \times \Lambda \rightarrow \mathbb{R}$ is bounded and Lipschitz continuous with uniform Lipschitz constant L and $g : \Lambda \rightarrow \mathbb{R}^\ell$ is locally Lipschitz.*
- *$\pi_1(W_\epsilon, W) = \mathcal{O}(\epsilon^{1/2-\delta})$ for all $\delta > 0$ where W is a $\mathcal{N}(0, \sigma^2)$ Brownian motion.*
- *For every $p \geq 1$ there exists a universal constant $C_p > 0$ such that for any $n \geq 1$ and $w : \Lambda \rightarrow \mathbb{R}$ Hölder continuous with $\int_\Lambda w d\mu = 0$ we have*

$$\left\| \int_0^n w \circ \phi_t dt \right\|_p \leq C_p \|w\|_{\text{Lip}} n^{1/2}.$$

Let us begin by observing the following:

Proposition 8.4. *The slow equation (8.7) can be written as*

$$x_\epsilon(t) = \xi + W_\epsilon(t) + \int_0^t a(x_\epsilon(s), y_\epsilon(s)) ds.$$

Proof: Consider the integral form of (8.7),

$$x_\epsilon(t) = \xi + \int_0^t a(x_\epsilon(s), y_\epsilon(s)) ds + \int_0^t \epsilon^{-1} v(y_\epsilon(s)) ds. \quad (8.11)$$

Observe that by change of variables we have $y_\epsilon(t) = y_1(t\epsilon^{-2}) = \phi_{t\epsilon^{-2}}$ and therefore

$$\int_0^t \epsilon^{-1} v(y_\epsilon(s)) ds = \epsilon^{-1} \int_0^t v \circ \phi_{s\epsilon^{-2}} ds = \epsilon \int_0^{t\epsilon^{-2}} v \circ \phi_s ds = W_\epsilon(t). \quad (8.12)$$

Substituting (8.12) in (8.11) proves the claim. ■

Proposition 8.5. *Assume that $h : \mathbb{R} \rightarrow \mathbb{R}$ is a Lipschitz continuous function with $\text{Lip}(h) = \lambda$. Let $\xi \in \mathbb{R}$ and $f \in \mathcal{C}([t_0 - \epsilon, t_0 + \epsilon])$ where $t_0 \in \mathbb{R}$, $\epsilon < \frac{1}{2\lambda}$. There exists a unique $u \in \mathcal{C}([t_0 - \epsilon, t_0 + \epsilon])$ satisfying*

$$u(t) = \xi + f(t) + \int_{t_0}^t h(u(s)) ds \quad \text{for all } t \in [t_0 - \epsilon, t_0 + \epsilon]. \quad (8.13)$$

Proof: Consider $\mathcal{F} : \mathcal{C}([t_0 - \epsilon, t_0 + \epsilon]) \rightarrow \mathcal{C}([t_0 - \epsilon, t_0 + \epsilon])$ where

$$\mathcal{F}(u)(t) = \xi + f(t) + \int_{t_0}^t h(u(s)) ds.$$

If $|t - t_0| < \epsilon$ then

$$\begin{aligned} |\mathcal{F}(u)(t) - \mathcal{F}(v)(t)| &\leq \left| \int_{t_0}^t (h(u(s)) - h(v(s))) ds \right| \leq \int_{t_0}^t |h(u(s)) - h(v(s))| ds \\ &\leq \lambda \int_{t_0}^t |u(s) - v(s)| ds \leq \lambda \int_{t_0}^t \|u - v\|_{\infty} ds \\ &= \lambda(t - t_0) \|u - v\|_{\infty} < \epsilon \lambda \|u - v\|_{\infty}. \end{aligned}$$

For $\epsilon < 1/2\lambda$ we have

$$\|\mathcal{F}(u)(t) - \mathcal{F}(v)(t)\|_{\infty} \leq \frac{1}{2} \|u - v\|_{\infty}$$

so that \mathcal{F} is a contraction mapping on $\mathcal{C}([t_0 - \epsilon, t_0 + \epsilon])$, a complete metric space. It then follows by the Banach fixed-point Theorem that there exists a unique $u \in \mathcal{C}([t_0 - \epsilon, t_0 + \epsilon])$ satisfying (8.13). ■

Recall the following inequality e.g. [63]:

Lemma 8.6. (Gronwall's Lemma) *If $u : [0, 1] \rightarrow [0, \infty)$ satisfies*

$$u(t) \leq C + K \int_0^t u(s) ds \quad \text{for all } t \in [0, 1]$$

where C, K are positive constants then $u(t) \leq Ce^{Kt}$ for all $t \in [0, 1]$.

Proposition 8.7. *Let $f_1, f_2 \in \mathcal{C}([0, 1])$ and assume that $h : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous with $\text{Lip}(h) = \lambda$. Also suppose that $u_1, u_2 \in \mathcal{C}([0, 1])$ satisfy*

$$u_i(t) = \xi + f_i(t) + \int_0^t h(u_i(s)) ds \quad \text{for all } t \in [0, 1], i = 1, 2.$$

Then $\|u_1 - u_2\|_\infty \leq e^\lambda \|f_1 - f_2\|_\infty$.

Proof: Observe that

$$\begin{aligned} |u_1(t) - u_2(t)| &\leq |f_1(t) - f_2(t)| + \left| \int_0^t (h(u_1(s)) - h(u_2(s))) ds \right| \\ &\leq |f_1(t) - f_2(t)| + \int_0^t |h(u_1(s)) - h(u_2(s))| ds \\ &\leq \|f_1 - f_2\|_\infty + \lambda \int_0^t |u_1(s) - u_2(s)| ds. \end{aligned}$$

It follows from Lemma 8.6 that

$$|u_1(t) - u_2(t)| \leq e^{\lambda t} \|f_1 - f_2\|_\infty \quad \text{for all } t \in [0, 1]$$

and we conclude that

$$\|u_1 - u_2\|_\infty \leq e^\lambda \|f_1 - f_2\|_\infty.$$

■

Theorem 8.8. *Assume that $h : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous with $\text{Lip}(h) = \lambda$. Define $\mathcal{G} : \mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ by $\mathcal{G}(f) = u$ where*

$$u(t) = \xi + f(t) + \int_0^t h(u(s)) ds \quad \text{for all } t \in [0, 1]. \quad (8.14)$$

Then \mathcal{G} is well-defined and Lipschitz continuous with constant e^λ .

Proof: We obtain global solutions to (8.14) by extending the local solutions, Proposition 8.5, from the interval $[t_0, t_0 + \epsilon]$ to the interval $[0, 1]$. Let $\epsilon = 1/2\lambda$

and, for $t_0 = 0$, let $I_1 = [0, \epsilon]$. It follows from Proposition 8.5 that there exists a unique and continuous $u_1 : I_1 \rightarrow \mathbb{R}$ such that

$$u_1(t) = \xi + f(t) + \int_0^t h(u(s)) ds \quad \text{for all } t \in I_1.$$

Next, take $I_2 = [\epsilon, 2\epsilon]$, that is $t_0 = \epsilon$. Then, by Proposition 8.5 there exists a unique and continuous $u_2(t) = u_1(\epsilon) - f(\epsilon) + f(t) + \int_\epsilon^t h(u(s)) ds$ for all $t \in I_2$. But then we have

$$\begin{aligned} u_2(t) &= \xi + f(\epsilon) - f(\epsilon) + f(t) + \int_0^\epsilon h(u(s)) ds + \int_\epsilon^t h(u(s)) ds \\ &= \xi + f(t) + \int_0^t h(u(s)) ds \quad \text{for all } t \in [0, 2\epsilon]. \end{aligned} \quad (8.15)$$

Take $n \in \mathbb{N}$ such that $(n-1)\epsilon < 1 \leq n\epsilon$. Then by Proposition 8.5 we find unique and continuous solutions u_3, \dots, u_{n-1} in the intervals $I_3, \dots, I_{n-1} = [(n-2)\epsilon, (n-1)\epsilon]$ respectively and therefore, by proceeding as in (8.15), in the interval $[0, (n-1)\epsilon]$ as well. That is, there exist unique and continuous u_3, \dots, u_{n-1} where

$$u_{n-1}(t) = u_{n-2}((n-2)\epsilon) - f((n-2)\epsilon) + f(t) + \int_{(n-2)\epsilon}^t h(u(s)) ds \quad \text{for all } t \in I_{n-1}$$

which extends to

$$u_{n-1}(t) = \xi + f(t) + \int_0^t h(u(s)) ds \quad \text{for all } t \in [0, (n-1)\epsilon].$$

Finally, we notice that $I = [0, 1] = \bigcup_{j=1}^n I_j \setminus (1, n\epsilon]$. We consider the interval $I_n = [(n-1)\epsilon, n\epsilon] \supseteq [(n-1)\epsilon, 1]$. To deal with the interval $(1, n\epsilon]$ we set $f(t) = f(1) - \int_1^t h(u(s)) ds$ for $t \geq 1$ so that the unique, continuous solution

$$u_n(t) = u_{n-1}((n-1)\epsilon) + f(t) - f((n-1)\epsilon) + \int_{(n-1)\epsilon}^t h(u(s)) ds$$

for all $t \in [(n-1)\epsilon, n\epsilon]$ extends to the interval $[0, 1]$ i.e.

$$u(t) = \xi + f(t) + \int_0^t h(u(s)) ds \quad \text{for all } t \in [0, 1].$$

Let $f \in \mathcal{C}([0, 1])$ and assume that there exist u, v both satisfying (8.14). Then by Proposition 8.7 we have that

$$|u(t) - v(t)| \leq (e^\lambda)0 \quad \text{for all } t \in [0, 1]$$

and we conclude that $u = v$. We have therefore that $\mathcal{G}(f) = u$ is well-defined. Lastly, note that by Proposition 8.7 we have

$$\|\mathcal{G}(f_1) - \mathcal{G}(f_2)\|_\infty = \|u_1 - u_2\|_\infty \leq e^\lambda \|f_1 - f_2\|_\infty$$

giving that $\text{Lip}(\mathcal{G}) = e^\lambda$. ■

We consider now the following special case of Theorem 8.2 (that gives a better error rate than in the general case).

Proposition 8.9. *Consider the fast-slow system of ODEs (8.7), (8.8) and denote by (x_ϵ, y_ϵ) its unique solution. Let Assumptions 8.3 hold and suppose in addition that $a(x, y) \equiv \bar{a}(x)$. Then $\pi_1(x_\epsilon, X) = \mathcal{O}(\epsilon^{1/2-\delta})$ for all $\delta > 0$.*

Proof: By Proposition 8.4 the slow equation is

$$x_\epsilon(t) = \xi + W_\epsilon(t) + \int_0^t \bar{a}(x_\epsilon(s)) ds.$$

By Theorem 8.8 there exists a well-defined, Lipschitz continuous mapping $\mathcal{G} : \mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ given by $\mathcal{G}(f) = u$ where

$$u(t) = \xi + f(t) + \int_0^t \bar{a}(u(s)) ds \quad \text{for all } t \in [0, 1].$$

Therefore we have that $\mathcal{G}(W_\epsilon) = x_\epsilon$. Also $\mathcal{G}(W) = X$ where $X = \xi + W + \int_0^t \bar{a}(X(s)) ds$. That is to say, $dX = dW + \bar{a}(X)dt$, $X(0) = \xi$. Lastly, it follows from Theorem 7.9 and Assumptions 8.3 that for all $\delta > 0$ there exists a real

constant $C > 0$ such that

$$\pi_1(x_\epsilon, X) = \pi_1(\mathcal{G}(W_\epsilon), \mathcal{G}(W)) \leq C\pi_1(W_\epsilon, W) \leq C\epsilon^{1/2-\delta}.$$

■

Building up to the proof of Theorem 8.2 we impose a boundedness condition for x_ϵ . We prove the following:

Theorem 8.10. *Consider the fast-slow system (8.7), (8.8) and the SDE (8.10). Let Assumptions 8.3 hold and assume in addition that there exists $Q_0 > 0$ such that $|x_\epsilon(t)| \leq Q_0$ for all $\epsilon > 0$ and all $t \in [0, 1]$. Then $\pi_1(x_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$.*

Proof: First let us define, following the argument in [55] that generalises the method of proof described in chapter 18 of [63],

$$Z_\epsilon(t) = \int_0^t \tilde{a}(x_\epsilon(s), y_\epsilon(s)) ds$$

where $\tilde{a}(x, y) = a(x, y) - \bar{a}(x) = a(x, y) - \int_\Omega a(x, y) d\mu(y)$. Note that $\|\tilde{a}\|_\infty \leq 2\|a\|_\infty$. We have

$$x_\epsilon(t) = \xi + W_\epsilon(t) + Z_\epsilon(t) + \int_0^t \bar{a}(x_\epsilon(s)) ds.$$

By Theorem 8.8 there exists a well-defined, Lipschitz continuous mapping $\mathcal{G} : \mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ such that $x_\epsilon = \mathcal{G}(W_\epsilon + Z_\epsilon)$.

Recall that $x_\epsilon(t) = \xi + \int_0^t a(x_\epsilon(s), y_\epsilon(s)) ds + \epsilon^{-1} \int_0^t v(y_\epsilon(s)) ds$ and observe that for $t_2 \geq t_1$ we have

$$\begin{aligned} |x_\epsilon(t_1) - x_\epsilon(t_2)| &\leq \left| \int_{t_1}^{t_2} a(x_\epsilon(s), y_\epsilon(s)) ds \right| + \left| \epsilon^{-1} \int_{t_1}^{t_2} v(y_\epsilon(s)) ds \right| \\ &\leq (\|a\|_\infty + \epsilon^{-1} \|v\|_\infty) (t_2 - t_1). \end{aligned} \tag{8.16}$$

Also note that

$$|Z_\epsilon(t) - Z_\epsilon([t\epsilon^{-k}]\epsilon^k)| \leq (t - [t\epsilon^{-k}]\epsilon^k) \|\tilde{a}\|_\infty \leq 2\epsilon^k \|a\|_\infty \quad (8.17)$$

where we take $k \in [1, 2]$. Moreover,

$$\begin{aligned} \int_{n\epsilon^k}^{(n+1)\epsilon^k} (\tilde{a}(x_\epsilon(s), y_\epsilon(s)) - \tilde{a}(x_\epsilon(n\epsilon^k), y_\epsilon(s))) ds &\leq \epsilon^k \text{Lip}(\tilde{a}) \sup_s |x_\epsilon(s) - x_\epsilon(n\epsilon^k)| \\ &\stackrel{(8.16)}{\leq} \epsilon^k \text{Lip}(\tilde{a}) (\|a\|_\infty + \epsilon^{-1} \|v\|_\infty) \sup_s (s - n\epsilon^k) = \mathcal{O}(\epsilon^{2k-1}). \end{aligned} \quad (8.18)$$

Note that $y_\epsilon(t) = y_1(t\epsilon^{-2})$. Therefore we have

$$\begin{aligned} Z_\epsilon(t) &= Z_\epsilon([t\epsilon^{-k}]\epsilon^k) + \mathcal{O}(\epsilon^k) \\ &= \sum_{n=0}^{[t\epsilon^{-k}]-1} \int_{n\epsilon^k}^{(n+1)\epsilon^k} \tilde{a}(x_\epsilon(s), y_\epsilon(s)) ds + \mathcal{O}(\epsilon^k) \\ &= \sum_{n=0}^{[t\epsilon^{-k}]-1} \int_{n\epsilon^k}^{(n+1)\epsilon^k} \tilde{a}(x_\epsilon(n\epsilon^k), y_\epsilon(s)) ds + \mathcal{O}(\epsilon^{k-1}) \\ &= \sum_{n=0}^{[t\epsilon^{-k}]-1} \epsilon^2 \int_{n\epsilon^{k-2}}^{(n+1)\epsilon^{k-2}} \tilde{a}(x_\epsilon(n\epsilon^k), y_1(s)) ds + \mathcal{O}(\epsilon^{k-1}) \\ &= \sum_{n=0}^{[t\epsilon^{-k}]-1} \epsilon^k J_\epsilon(n) + \mathcal{O}(\epsilon^{k-1}), \end{aligned}$$

where $J_\epsilon(n) = \epsilon^{2-k} \int_{n\epsilon^{k-2}}^{(n+1)\epsilon^{k-2}} \tilde{a}(x_\epsilon(n\epsilon^k), y_1(s)) ds$. The first equality above follows from (8.17) and the third equality from (8.18).

Hence we have

$$\max_{[0,1]} |Z_\epsilon| \leq \sum_{n=0}^{[\epsilon^{-k}]-1} \epsilon^k |J_\epsilon(n)| + \mathcal{O}(\epsilon^{k-1}).$$

For $u \in \mathbb{R}$ fixed we define

$$\tilde{J}_\epsilon(n, u) = \epsilon^{2-k} \int_{n\epsilon^{k-2}}^{(n+1)\epsilon^{k-2}} \tilde{a}(u, y_1(s)) ds = \epsilon^{2-k} \int_{n\epsilon^{k-2}}^{(n+1)\epsilon^{k-2}} A_u \circ \phi_s ds$$

where $A_u(y) = \tilde{a}(u, y)$. Since $\tilde{J}_\epsilon(n, u) = \tilde{J}_\epsilon(0, u) \circ \phi_{n\epsilon^{k-2}}$ we have that

$$\int_{\Lambda} \left| \tilde{J}_\epsilon(n, u) \right|^p d\mu = \int_{\Lambda} \left| \tilde{J}_\epsilon(0, u) \right|^p d\mu$$

for all $p \geq 1$.

Consider a grid S on $[-Q_0, Q_0] \subset \mathbb{R}$ such that for all $x_\epsilon(n\epsilon^k)$ there exists a $u \in S$ with $|u - x_\epsilon(n\epsilon^k)| \leq 2Q_0/|S|$. Then

$$|\tilde{a}(u, y_1(s)) - \tilde{a}(x_\epsilon(n\epsilon^k), y_1(s))| \leq \text{Lip}(\tilde{a}) \|u - x_\epsilon(n\epsilon^k)\|_\infty \leq \frac{4LQ_0}{|S|}. \quad (8.19)$$

Since

$$|J_\epsilon(n)| \leq |\tilde{J}_\epsilon(n, u)| + |J_\epsilon(n) - \tilde{J}_\epsilon(n, u)|$$

we have from (8.19) that for all $n \geq 0$ and $\epsilon > 0$

$$|J_\epsilon(n)| \leq \sum_{u \in S} |\tilde{J}_\epsilon(n, u)| + \frac{4LQ_0}{|S|}.$$

Hence by Minkowski's inequality we have for all $p \geq 1$

$$\begin{aligned} \left\| \max_{[0,1]} |Z_\epsilon| \right\|_p &\leq \sum_{n=0}^{[\epsilon^{-k}]-1} \epsilon^k \left\| \sum_{u \in S} |\tilde{J}_\epsilon(n, u)| \right\|_p + \frac{4LQ_0}{|S|} + \mathcal{O}(\epsilon^{k-1}) \\ &= \sum_{n=0}^{[\epsilon^{-k}]-1} \epsilon^k \left\| \sum_{u \in S} |\tilde{J}_\epsilon(0, u)| \right\|_p + \frac{4LQ_0}{|S|} + \mathcal{O}(\epsilon^{k-1}) \\ &= \sum_{u \in S} \left\| \tilde{J}_\epsilon(0, u) \right\|_p + \frac{4LQ_0}{|S|} + \mathcal{O}(\epsilon^{k-1}). \end{aligned} \quad (8.20)$$

Now we have by definition of \tilde{a} that A_u is mean-zero. Moreover, A_u is Lipschitz continuous since a is. Thus by Assumptions 8.3 we have

$$\left\| \tilde{J}_\epsilon(0, u) \right\|_p = \epsilon^{2-k} \left\| \int_0^{\epsilon^{k-2}} A_u \circ \phi_s ds \right\|_p \leq C_p \text{Lip}(\tilde{a}) \epsilon^{(2-k)/2}.$$

Hence there exists a constant $C > 0$ such that

$$\left\| \max_{[0,1]} |Z_\epsilon| \right\|_p \leq C \left(|S| \epsilon^{(2-k)/2} + \frac{1}{|S|} + \epsilon^{k-1} \right). \quad (8.21)$$

It follows that the optimal bound for Z_ϵ is attained at $S = S_\epsilon = \mathcal{O}(\epsilon^{-1/5})$ and $k = 6/5$ giving that

$$\left\| \max_{[0,1]} |Z_\epsilon| \right\|_p \leq C \epsilon^{1/5}.$$

Therefore from Proposition 6.1 we obtain that there exists a real constant $C > 0$ such that $\pi_1(W_\epsilon + Z_\epsilon, W_\epsilon)$ is bounded by $C \epsilon^{\frac{1}{5} \frac{p}{p+1}}$. Observe that by Theorem 7.9 we have

$$\pi_1(x_\epsilon, X) = \pi_1(\mathcal{G}(W_\epsilon + Z_\epsilon), \mathcal{G}(W)) \leq (e^L \vee 1) \pi_1(W_\epsilon + Z_\epsilon, W).$$

Finally note that for all $\delta' > 0$ and for all $p \geq 1$

$$\begin{aligned} \pi_1(W_\epsilon + Z_\epsilon, W) &\leq \pi_1(W_\epsilon, W) + \pi_1(W_\epsilon + Z_\epsilon, W_\epsilon) \\ &= \mathcal{O}(\epsilon^{\frac{1}{2}-\delta'}) + \mathcal{O}(\epsilon^{\frac{1}{5} \frac{p}{p+1}}) = \mathcal{O}(\epsilon^{\frac{1}{5}-\delta}) \end{aligned}$$

where $1/\delta = 5(p+1)$. Thus we conclude that $\pi_1(x_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$ as required. \blacksquare

Relaxing the boundedness condition for x_ϵ in Theorem 8.10 brings us into the setting of Theorem 8.2. First let us recall a standard result.

Lemma 8.11. *Let W be a Brownian motion with variance σ^2 . Then, for $c > 0$ we have*

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq c] \leq \frac{2\sqrt{2}\sigma}{c\sqrt{\pi}} e^{-\frac{c^2}{2\sigma^2}}.$$

Proof: Recall that the reflection principle for a Brownian motion gives that for any $c > 0$

$$\mathbb{P}[\max_{0 \leq s \leq t} W(s) \geq c] = 2\mathbb{P}[W(t) \geq c].$$

By symmetry of a Brownian motion we have

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq c] \leq 2\mathbb{P}[\max_{s \in [0,1]} W(s) \geq c] = 4\mathbb{P}[W(1) \geq c].$$

Here $W(1) =_d \mathcal{N}(0, \sigma^2)$ i.e. its distribution function is given by

$$\mathbb{P}[W(1) \leq y] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y e^{\frac{-s^2}{2\sigma^2}} ds.$$

Since $\frac{d}{ds} \left(e^{\frac{-s^2}{2\sigma^2}} \right) = -\frac{s}{\sigma^2} e^{\frac{-s^2}{2\sigma^2}}$ integration by parts gives

$$\int_y^\infty e^{\frac{-s^2}{2\sigma^2}} ds = \int_y^\infty -\frac{\sigma^2}{s} \left(-\frac{s}{\sigma^2} e^{\frac{-s^2}{2\sigma^2}} \right) ds = \left[-\frac{\sigma^2}{s} e^{\frac{-s^2}{2\sigma^2}} \right]_y^\infty - \int_y^\infty \frac{\sigma^2}{s^2} e^{\frac{-s^2}{2\sigma^2}} ds.$$

The rightmost integral above is strictly positive and therefore

$$\int_y^\infty e^{\frac{-s^2}{2\sigma^2}} ds < \sigma^2 y^{-1} e^{\frac{-y^2}{2\sigma^2}}.$$

It follows that

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq c] \leq \frac{2\sqrt{2}\sigma}{c\sqrt{\pi}} e^{-\frac{c^2}{2\sigma^2}}.$$

■

We are now in position to prove Theorem 8.2.

Proof of Theorem 8.2: For $Q > 0$ we write $Z_\epsilon = Z_\epsilon^{Q,1} + Z_\epsilon^{Q,2}$ where

$$Z_\epsilon^{Q,1}(t) = Z_\epsilon(t) \mathbf{1}_{B_\epsilon(Q)}, \quad Z_\epsilon^{Q,2}(t) = Z_\epsilon(t) \mathbf{1}_{B_\epsilon(Q)^c}, \quad B_\epsilon(Q) = \left\{ \max_{[0,1]} |x_\epsilon| \leq Q \right\}.$$

Since $|x_\epsilon - W_\epsilon| \leq \xi + \|a\|_\infty$ we have for $Q > 2(\xi + \|a\|_\infty)$ that

$$\begin{aligned} \mu[\max_{[0,1]} |Z_\epsilon^{Q,2}| > 0] &\leq \mu[\max_{[0,1]} |x_\epsilon| > Q] \leq \mu[\max_{[0,1]} |W_\epsilon| > Q/2] \\ &= \mathbb{P}[\max_{[0,1]} |W| > Q/4] + \left(\mu[\max_{[0,1]} |W_\epsilon| > Q/2] - \mathbb{P}[\max_{[0,1]} |W| > Q/4] \right). \end{aligned} \quad (8.22)$$

Note that by Lemma 8.11 we have

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq Q/4] \leq \frac{8\sqrt{2}\sigma}{Q\sqrt{\pi}} e^{-\frac{Q^2}{32\sigma^2}}. \quad (8.23)$$

Let $Q = Q_\epsilon = \sqrt{-q \log \epsilon}$ where $q > 0$. Observe that $Q_\epsilon \rightarrow \infty$ as $\epsilon \rightarrow 0$ for any positive constant q . In particular, assume that $Q_\epsilon > 8\sqrt{2}\sigma/\sqrt{\pi}$ and choose $q = 32\sigma^2$. Then (8.23) gives

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq Q_\epsilon/4] \leq \frac{8\sqrt{2}\sigma}{Q_\epsilon\sqrt{\pi}} e^{(32\sigma^2/32\sigma^2) \log \epsilon} < \frac{Q_\epsilon}{Q_\epsilon} e^{\log \epsilon} = \epsilon. \quad (8.24)$$

Now by assumption we have that $\pi_1(W_\epsilon, W) \leq C\epsilon^{1/2-\delta}$ for some $C > 0$. Let $d = 2C\epsilon^{1/2-\delta}$. Then $\mu[W_\epsilon \in A] \leq \mu[W_\epsilon \in A^d] + d$ for all Borel sets A . In particular, let $A = \{u \in \mathcal{C}([0,1]) : \max_{[0,1]} |u| \geq \frac{Q_\epsilon}{2}\}$. Note that if $W \in A^d$ then there exists some $u \in A$ such that $\max_{[0,1]} |W - u| \leq d$. It follows from the (reverse) triangle inequality that $|\max_{[0,1]} |W| - \max_{[0,1]} |u|| \leq \max_{[0,1]} |W - u| \leq d$ and in particular $\max_{[0,1]} |W| - \max_{[0,1]} |u| \geq -d$. Hence $\max_{[0,1]} |W| \geq \max_{[0,1]} |u| - d \geq Q_\epsilon/2 - d$. Thus we have

$$\begin{aligned} \mu[\max_{[0,1]} |W_\epsilon| > Q_\epsilon/2] &= \mu[W_\epsilon \in A] \leq \mathbb{P}[W \in A^d] + d \\ &\leq \mathbb{P}[\max_{[0,1]} |W| > Q_\epsilon/2 - d] + d \leq \mathbb{P}[\max_{[0,1]} |W| > Q_\epsilon/4] + d \end{aligned}$$

where the last inequality follows from $Q_\epsilon > 4d$. Rearranging we obtain

$$\mu[\max_{[0,1]} |W_\epsilon| > Q_\epsilon/2] - \mathbb{P}[\max_{[0,1]} |W| > Q_\epsilon/4] \leq d. \quad (8.25)$$

Putting together equations (8.24) and (8.25) into (8.22) gives

$$\mu[\max_{[0,1]} |Z_\epsilon^{Q_\epsilon, 2}| > 0] < \epsilon + d = \epsilon + 2C\epsilon^{1/2-\delta} = \mathcal{O}(\epsilon^{1/2-\delta}).$$

Recall next that $Z_\epsilon^{Q_\epsilon, 1} = Z_\epsilon \mathbb{1}_{B_\epsilon(Q_\epsilon)}$. Observe that $\mathbb{1}_{B_\epsilon(Q_\epsilon)}$ brings us directly in the setting of Theorem 8.10 with x_ϵ bounded by Q_ϵ . In particular let S_ϵ be a grid on $[-Q_\epsilon, Q_\epsilon] \subset \mathbb{R}$ such that for all $x_\epsilon(n\epsilon^k)$ there exists a $u \in S_\epsilon$ with

$|u - x_\epsilon(n\epsilon^k)| \leq 2Q_\epsilon/|S_\epsilon|$. Then the estimate in (8.20) holds true with Q_0 being replaced by Q_ϵ and S by S_ϵ . The \mathcal{L}^p estimate in the hypothesis of Theorem 8.10 can now be applied and (8.21) is changed to

$$\left\| \max_{[0,1]} |Z_\epsilon^{Q_\epsilon,1}| \right\|_p \leq C \left(|S_\epsilon| \epsilon^{(2-k)/2} + \frac{Q_\epsilon}{|S_\epsilon|} + \epsilon^{k-1} \right).$$

for some $C > 0$. We obtain the optimal bound for $Z_\epsilon^{Q_\epsilon,1}$ at $k = 6/5$ and $|S_\epsilon| \approx \epsilon^{-1/5}$ giving that

$$\left\| \max_{[0,1]} |Z_\epsilon^{Q_\epsilon,1}| \right\|_p \leq C \epsilon^{1/5} \sqrt{-\log \epsilon} = \mathcal{O}(\epsilon^{1/5-\delta'})$$

where $\delta' > 0$ is arbitrarily small. Since

$$\left\| \max_{t \in [0,1]} |(W_\epsilon + Z_\epsilon^{Q_\epsilon,1} + Z_\epsilon^{Q_\epsilon,2}) - W_\epsilon| \right\|_p \leq \left\| \max_{t \in [0,1]} |Z_\epsilon^{Q_\epsilon,1}| \right\|_p + \left\| \max_{t \in [0,1]} |Z_\epsilon^{Q_\epsilon,2}| \right\|_p$$

we deduce from Proposition 7.4 that there exists a real constant $C > 0$ such that $\pi_1(W_\epsilon + Z_\epsilon, W_\epsilon)$ is bounded by $C \epsilon^{(\frac{1}{5}-\delta')\frac{p}{p+1}}$. We conclude, as at the end of the proof of Theorem 8.10, that $\pi_1(x_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$. ■

8.4 Rates of Convergence in the Multiplicative Noise case for Flows

We now investigate convergence rates for flows in the presence of multiplicative noise.

Theorem 8.12. (Main Theorem for Flows) *Consider the fast-slow system (8.9), (8.8). Let Assumptions 8.3 hold. Assume that $b : \mathbb{R} \rightarrow \mathbb{R}$ is uniformly Lipschitz and bounded and also that $1/b$ is bounded. Then $\pi_1(x_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$ where X is the unique solution to the Stratonovich SDE*

$$dX = \bar{a}(X)dt + b(X) \circ dW, \quad x_\epsilon(0) = \xi.$$

Proof: Let us write $z_\epsilon = \psi(x_\epsilon)$ where $\psi' = 1/b$. Observe that

$$\dot{z}_\epsilon = \psi'(x_\epsilon)\dot{x}_\epsilon = b(x_\epsilon)^{-1}(a(x_\epsilon, y_\epsilon) + b(x_\epsilon)\dot{W}_\epsilon) = b(x_\epsilon)^{-1}a(x_\epsilon, y_\epsilon) + \dot{W}_\epsilon.$$

Denote $\alpha(z, y) = \psi'(\psi^{-1}(z))a(\psi^{-1}(z), y)$. Then we obtain a slow ODE

$$\dot{z}_\epsilon = \alpha(z_\epsilon, y_\epsilon) + \dot{W}_\epsilon, \quad z_\epsilon(0) = \psi(\xi).$$

Write

$$dZ = \bar{\alpha}(Z)dt + dW$$

where $\bar{\alpha}(z) = \int_{\Omega} \alpha(z, y) d\mu(y) = \psi'(\psi^{-1}(z))\bar{a}(\psi^{-1}(z))$.

Let us observe that by the Inverse Function Theorem we have

$$\frac{1}{(\psi')}(\psi^{-1}(z)) = (\psi^{-1})'(z).$$

By assumption we have that $1/\psi' = b$ is bounded and therefore we deduce by the Mean Value Theorem that ψ^{-1} is Lipschitz continuous. Also, since ψ' is bounded and $1/\psi'$ is uniformly Lipschitz it follows that ψ' is uniformly Lipschitz. Thus we obtain that α is uniformly Lipschitz continuous, being the product and composition of uniformly Lipschitz continuous functions.

We can therefore apply Theorem 8.2 to obtain that $z_\epsilon \rightarrow_w Z$ with error rate $\pi_1(z_\epsilon, Z) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$. Since the chain rule in Stratonovich's Theory satisfies the usual laws of ordinary deterministic calculus we obtain that the limiting process $X = \psi^{-1}(Z)$ is given by

$$\begin{aligned} dX &= (\psi^{-1})'(Z) \circ dZ = \psi'(X)^{-1}(\bar{\alpha}(Z)dt + \circ dW) \\ &= \psi'(X)^{-1}(\psi'(X)\bar{a}(X)dt + \circ dW) = \bar{a}(X)dt + b(X) \circ dW, \end{aligned}$$

where the second equality follows from the derivative of the inverse function formula. Consider now a functional $\mathcal{H} : \mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ given by $\mathcal{H}(u) =$

$\vartheta \circ u$ where $\vartheta : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz-continuous. Then

$$\sup |\mathcal{H}(u) - \mathcal{H}(v)| = \sup |\vartheta \circ u - \vartheta \circ v| \leq \text{Lip}(\vartheta) \sup |u - v|$$

so that $\text{Lip}(\mathcal{H}) \leq \text{Lip}(\vartheta)$. Letting $\vartheta = \psi^{-1}$ we can apply Theorem 7.9 (Lipschitz Mapping Theorem) to conclude that for all $\delta > 0$

$$\pi_1(x_\epsilon, X) = \pi_1(\mathcal{H}(z_\epsilon), \mathcal{H}(Z)) \leq (\text{Lip}(\psi^{-1}) \vee 1) \pi_1(z_\epsilon, Z) = \mathcal{O}(\epsilon^{1/5-\delta}).$$

■

Chapter 9

Fast-Slow systems in Discrete Time

9.1 Setting and Assumptions

In ergodic theory, it is well known (e.g. [60]) that proving statistical limit laws is easier for discrete time dynamical systems than for continuous time systems. However, for fast-slow systems the situation is reversed: given good understanding of the fast dynamics, the homogenization problem is easier for continuous time. Hence we have proved convergence rates for homogenization for continuous time in chapter 8 in which the strong statistical properties satisfied by the fast dynamics were assumed. In this chapter we obtain rates of convergence for fast-slow systems in the discrete-time setting under the assumption that the fast dynamics is uniformly expanding – this time the required assumptions on the fast dynamics follow from the results under the General Framework **(H)** in the earlier chapters in this thesis (in particular, martingale approximations, Burkholder estimates, and error rates in the Weak Invariance Principle). As we shall see, the error rates we obtain in homogenization for deterministic maps in the discrete-time setting replicate the ones obtained in chapter 8 for flows.

We first consider the case where we have additive noise. Let

$$x_\epsilon(n+1) = x_\epsilon(n) + \epsilon v(y(n)) + \epsilon^2 a(x_\epsilon(n), y(n), \epsilon), \quad x_\epsilon(0) = \xi. \quad (9.1)$$

Here $x_\epsilon(n) \in \mathbb{R}$ and the fast variables $y(n)$ are generated by a map $T : \Lambda \rightarrow \Lambda$ with compact attractor Λ and ergodic invariant measure μ . Λ is assumed to satisfy certain “mild chaoticity” conditions stated precisely in Assumptions 9.2 below. Given $y_0 = \eta \in \Lambda$, the fast variables $y(n)$, $n \geq 0$ are defined by setting $y(n+1) = T(y(n))$. Later on we will focus on the multiplicative noise case by considering

$$x_\epsilon(n+1) = x_\epsilon(n) + \epsilon b(x_\epsilon(n))v(y(n)) + \epsilon^2 a(x_\epsilon(n), y(n), \epsilon) \quad (9.2)$$

where $b : \mathbb{R} \rightarrow \mathbb{R}$ satisfies some regularity conditions.

Define $\hat{x}_\epsilon(t) = x_\epsilon(t\epsilon^{-2})$ for $t = 0, \epsilon^2, 2\epsilon^2, \dots$ and linearly interpolating obtain $\hat{x} \in \mathcal{C}([0, 1])$.

Also set $\bar{a}(x) = \int_\Lambda a(x, y, 0) d\mu(y)$ and let X denote solutions of the SDE of the form

$$dX = dW + \bar{a}(X)dt, \quad X(0) = \xi. \quad (9.3)$$

Here $W \sim \mathcal{N}(0, \sigma^2)$ is a Brownian motion and the variance parameter σ^2 corresponds to the given observable v as deduced in Proposition 4.4 i.e. $\sigma^2 = \int m^2 d\mu$.

The issue of weak convergence was answered by Gottwald & Melbourne in [32], Theorem 1.1 and Proposition 1.5 for the additive noise and the multiplicative noise case respectively. In the next section we obtain the rate of convergence \hat{x}_ϵ to X in $\mathcal{C}([0, 1])$. The multiplicative noise case will then be addressed in the final section with the limiting SDE being of Stratonovich type as shown in [32].

9.2 Rates of Convergence in the Additive Noise case for Maps

We prove the following:

Theorem 9.1. *Consider equation (9.1) and the SDE (9.3). Let Assumptions 9.2 below hold. Then $\pi_1(\hat{x}_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$.*

We make the following standing assumptions.

Assumptions 9.2. *Equation (9.1) satisfies the following:*

- *The observable $v : \Lambda \rightarrow \mathbb{R}$ is Hölder continuous and $\int v d\mu = 0$.*
- *The map $T : \Lambda \rightarrow \Lambda$ satisfies the General Framework (H).*
- *$\xi \in \mathbb{R}$ is fixed throughout and $\eta \in \Lambda$ is the sole source of randomness in the fast-slow system.*
- *$a : \mathbb{R} \times \Lambda \times \mathbb{R} \rightarrow \mathbb{R}$ is bounded and Lipschitz continuous with uniform Lipschitz constant L .*
- *There exists a real constant $C > 0$ such that*

$$|a(x, y, \epsilon) - a(x, y, 0)| \leq C\epsilon^{1/5} \quad (9.4)$$

for all $x \in \mathbb{R}$, $y \in \Lambda$ and $\epsilon \in (0, 1]$.

Let us first prove the following:

Lemma 9.3. *Consider equation (9.1) and let Assumptions 9.2 hold. Define $W_\epsilon(t) = \epsilon \sum_{j=0}^{t\epsilon^{-2}-1} v(y(j))$ for $t = 0, \epsilon^2, 2\epsilon^2, \dots$ and linearly interpolate to obtain $W_\epsilon \in \mathcal{C}([0, 1])$. Also let*

$$Z_\epsilon(t) = \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} [a(x_\epsilon(j), y(j), 0) - \bar{a}(x_\epsilon(j))]. \quad (9.5)$$

Then we have

$$\hat{x}_\epsilon(t) = \xi + W_\epsilon(t) + \hat{Z}_\epsilon(t) + \int_0^t a(\hat{x}_\epsilon(s)) ds$$

where $|\hat{Z}_\epsilon(t) - Z_\epsilon(t)| \leq C\epsilon^{1/5}$ for some real constant $C > 0$ and for all $\epsilon \in (0, 1]$.

Proof: First note that iterating (9.1) we can rewrite $x_\epsilon(n)$ as

$$x_\epsilon(n) = \xi + \epsilon \sum_{j=0}^{n-1} v(y(j)) + \epsilon^2 \sum_{j=0}^{n-1} a(x_\epsilon(j), y(j), \epsilon).$$

Introduce the piecewise constant function $\tilde{x}_\epsilon(t) = x_\epsilon([t\epsilon^{-2}])$. We have

$$\tilde{x}_\epsilon(t) = \xi + \epsilon \sum_{j=0}^{[t\epsilon^{-2}]-1} v(y(j)) + \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} a(x_\epsilon(j), y(j), \epsilon). \quad (9.6)$$

Observe that

$$\|\hat{x}_\epsilon(t) - \tilde{x}_\epsilon(t)\|_\infty \leq \epsilon \|v\|_\infty + \epsilon^2 \|a\|_\infty.$$

Hence by (9.6) we have

$$\hat{x}_\epsilon(t) = \tilde{x}_\epsilon(t) + \mathcal{O}(\epsilon) = \xi + \epsilon \sum_{j=0}^{[t\epsilon^{-2}]-1} v(y(j)) + \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} a(x_\epsilon(j), y(j), \epsilon) + \mathcal{O}(\epsilon). \quad (9.7)$$

Notice that

$$\left\| W_\epsilon - \epsilon \sum_{j=0}^{[t\epsilon^{-2}]-1} v(y(j)) \right\|_\infty \leq \epsilon \|v\|_\infty \quad (9.8)$$

and that by Assumptions 9.2, (9.4) we have

$$\left| \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} a(x_\epsilon(j), y(j), \epsilon) - \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} a(x_\epsilon(j), y(j), 0) \right| \leq C\epsilon^{1/5}. \quad (9.9)$$

Substituting (9.8) and (9.9) in (9.7) we have

$$\hat{x}_\epsilon(t) = \xi + W_\epsilon(t) + \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} a(x_\epsilon(j), y(j), 0) + \mathcal{O}(\epsilon^{1/5}). \quad (9.10)$$

Now it is clear that for $j \in \mathbb{N}$ we have $\hat{x}_\epsilon(\epsilon^2 j) = x_\epsilon(j)$. Writing

$$a(x_\epsilon(j), y(j), 0) = \bar{a}(\hat{x}_\epsilon(\epsilon^2 j)) + [a(x_\epsilon(j), y(j), 0) - \bar{a}(x_\epsilon(j))]$$

we have that

$$\epsilon^2 \sum_{j=0}^{\lfloor t\epsilon^{-2} \rfloor - 1} a(x_\epsilon(j), y(j), 0) = Z_\epsilon(t) + \epsilon^2 \sum_{j=0}^{\lfloor t\epsilon^{-2} \rfloor - 1} \bar{a}(\hat{x}_\epsilon(\epsilon^2 j)). \quad (9.11)$$

Observe that if $t = n\epsilon^2$ where $n \in \mathbb{N}$ then the rightmost term in (9.11) is the Riemann sum of a piecewise constant function and is precisely $\int_0^t \bar{a}(\hat{x}_\epsilon(j)) ds$. Otherwise, we have $t = n\epsilon^2 + r$ where $n \in \mathbb{N}$ and $0 < r < \epsilon^2$. Therefore we obtain that

$$\epsilon^2 \sum_{j=0}^{\lfloor t\epsilon^{-2} \rfloor - 1} \bar{a}(\hat{x}_\epsilon(\epsilon^2 j)) = \int_0^t \bar{a}(\hat{x}_\epsilon(j)) ds + \mathcal{O}(\epsilon^2).$$

Substituting in (9.10) we deduce that

$$\hat{x}_\epsilon(t) = \xi + W_\epsilon(t) + \hat{Z}_\epsilon(t) + \int_0^t \bar{a}(\hat{x}_\epsilon(s)) ds,$$

where $|\hat{Z}_\epsilon(t) - Z_\epsilon(t)| \leq C\epsilon^{1/5}$ for some real constant $C > 0$ and for all $\epsilon \in (0, 1]$. ■

We proceed by estimating Z_ϵ following the argument in [32], Appendix A combined with chapter 7.

Proposition 9.4. *Let Z_ϵ be as in (9.5) and Assumptions 9.2 hold. Then for all $p > 1$ we have*

$$\left\| \max_{[0,1]} Z_\epsilon \right\|_p = \mathcal{O}(\epsilon^{(\frac{1}{5}-\delta)}).$$

Proof: Firstly note that $\|\tilde{a}\|_\infty \leq 2\|a\|_\infty$ and $\text{Lip}(\tilde{a}) \leq 2\text{Lip}(a) = 2L$. Define $\tilde{a}(x, y) = a(x, y, 0) - \bar{a}(x)$ and rewrite (9.5) as

$$Z_\epsilon(t) = \epsilon^2 \sum_{j=0}^{\lfloor t\epsilon^{-2} \rfloor - 1} \tilde{a}(x_\epsilon(j), y(j)).$$

Let $N = \lfloor t/\epsilon^k \rfloor$ where $k \in [1, 2]$ and set $H_\epsilon = \epsilon^2 \sum_{\lfloor N\epsilon^{k-2} \rfloor \leq j < \lfloor t\epsilon^{-2} \rfloor} \tilde{a}(x_\epsilon(j), y(j))$. Then $Z_\epsilon(t) = Z_\epsilon(N\epsilon^k) + H_\epsilon$. Recall that for positive real numbers $x, y \in \mathbb{R}^+$

with $x \geq y$ we have that $[x] + [y] \leq [x + y] \leq [x] + [y] + 1$ and $[x - y] \leq [x] - [y] \leq [x - y] + 1$.

We have

$$\begin{aligned} |H_\epsilon| &\leq \epsilon^2([t\epsilon^{-2}] - [N\epsilon^{k-2}])\|\tilde{a}\|_\infty \leq 2\|a\|_\infty \epsilon^2(t\epsilon^{-2} - N\epsilon^{k-2} + 1) \\ &= 2\|a\|_\infty((t - [t\epsilon^{-k}]\epsilon^k + \epsilon^2) \leq 2\|a\|_\infty(t - (t\epsilon^{-k} - 1)\epsilon^k + \epsilon^2) \\ &\leq 2\|a\|_\infty(\epsilon^k + \epsilon^2) \leq 4\|a\|_\infty \epsilon^k \end{aligned}$$

since $k \in [1, 2]$.

Next, we estimate $Z_\epsilon(N\epsilon^k)$ as follows:

$$\begin{aligned} Z_\epsilon(N\epsilon^k) &= \epsilon^2 \sum_{n=0}^{N-1} \sum_{[n\epsilon^{k-2}] \leq j < [(n+1)\epsilon^{k-2}]} \tilde{a}(x_\epsilon(j), y(j)) \\ &= \epsilon^2 \sum_{n=0}^{N-1} \sum_{[n\epsilon^{k-2}] \leq j < [(n+1)\epsilon^{k-2}]} (\tilde{a}(x_\epsilon(j), y(j)) - \tilde{a}(x_\epsilon(n\epsilon^{k-2}), y(j))) \\ &\quad + \epsilon^2 \sum_{n=0}^{N-1} \sum_{[n\epsilon^{k-2}] \leq j < [(n+1)\epsilon^{k-2}]} \tilde{a}(x_\epsilon(n\epsilon^{k-2}), y(j)) \\ &= I_\epsilon + \Upsilon_\epsilon. \end{aligned}$$

Note that for $k \in [1, 2]$

$$[(n+1)\epsilon^{k-2}] - [n\epsilon^{k-2}] \leq (n+1)\epsilon^{k-2} - n\epsilon^{k-2} + 1 = \epsilon^{k-2} + 1 \leq 2\epsilon^{k-2}.$$

Observe that for $[n\epsilon^{k-2}] \leq j < [(n+1)\epsilon^{k-2}]$ we have

$$\begin{aligned} |x_\epsilon(j) - x_\epsilon(n\epsilon^{k-2})| &\leq (\epsilon\|v\|_\infty + \epsilon^2\|a\|_\infty)(j - [n\epsilon^{k-2}]) \\ &\leq (\epsilon\|v\|_\infty + \epsilon^2\|a\|_\infty)[(n+1)\epsilon^{k-2}] - [n\epsilon^{k-2}] \\ &\leq 2(\|v\|_\infty + \epsilon\|a\|_\infty)\epsilon^{k-1}. \end{aligned}$$

Hence for $\epsilon \in (0, 1)$ we obtain

$$\begin{aligned} |I_\epsilon| &\leq 2\epsilon^2 N([(n+1)\epsilon^{k-2}] - [n\epsilon^{k-2}]) \text{Lip}(\tilde{a})(\|v\|_\infty + \epsilon\|a\|_\infty)\epsilon^{k-1} \\ &\leq 4\epsilon^2 [t\epsilon^{-k}] \epsilon^{k-2} \text{Lip}(\tilde{a})(\|v\|_\infty + \epsilon\|a\|_\infty)\epsilon^{k-1} \\ &\leq 4t \text{Lip}(\tilde{a})(\|v\|_\infty + \|a\|_\infty)\epsilon^{k-1} \leq 8 \text{Lip}(a)(\|v\|_\infty + \|a\|_\infty)\epsilon^{k-1}. \end{aligned}$$

Consider next

$$\Upsilon_\epsilon = \epsilon^2 \sum_{n=0}^{N-1} \sum_{j=[n\epsilon^{k-2}]}^{[(n+1)\epsilon^{k-2}]-1} \tilde{a}(x_\epsilon(n\epsilon^{k-2}), y(j)) = \epsilon^k \sum_{n=0}^{N-1} J_\epsilon(n),$$

where

$$J_\epsilon(n) = \epsilon^{2-k} \sum_{j=[n\epsilon^{k-2}]}^{[(n+1)\epsilon^{k-2}]-1} \tilde{a}(x_\epsilon(n\epsilon^{k-2}), y(j)).$$

Note that

$$|\Upsilon_\epsilon| \leq \epsilon^k \sum_{n=0}^{[\epsilon^{-k}]-1} |J_\epsilon(n)|.$$

For $u \in \mathbb{R}$ fixed we define

$$\tilde{J}_\epsilon(n, u) = \epsilon^{2-k} \sum_{j=[n\epsilon^{k-2}]}^{[(n+1)\epsilon^{k-2}]-1} \tilde{a}(u, y(j)) = \epsilon^{2-k} \sum_{j=[n\epsilon^{k-2}]}^{[(n+1)\epsilon^{k-2}]-1} A_u \circ T^j$$

where $A_u(y) = \tilde{a}(u, y)$. Then

$$\tilde{J}_\epsilon(0, u) = \epsilon^{2-k} \sum_{j=0}^{[\epsilon^{k-2}]-1} A_u \circ T^j.$$

It is immediate that

$$\tilde{J}_\epsilon(0, u) \circ T^{[n\epsilon^{k-2}]} = \epsilon^{2-k} \sum_{j=[n\epsilon^{k-2}]}^{[n\epsilon^{k-2}]+[\epsilon^{k-2}]-1} A_u \circ T^j.$$

Since

$$[(n+1)\epsilon^{k-2}] - 2 \leq [n\epsilon^{k-2}] + [\epsilon^{k-2}] - 1 \leq [(n+1)\epsilon^{k-2}] - 1$$

we have that

$$\tilde{J}_\epsilon(0, u) \circ T^{[n\epsilon^{k-2}]} = \tilde{J}_\epsilon(n, u) - \beta\epsilon^{2-k}A_u \circ T^{[(n+1)\epsilon^{k-2}]-1}$$

where $\beta \in \{0, 1\}$. Since the map T is measure-preserving it follows that

$$\mathbb{E} \left| \tilde{J}_\epsilon(n, u) \right| \leq \mathbb{E} \left| \tilde{J}_\epsilon(0, u) \right| + \epsilon^{2-k} \mathbb{E} |A_u| \leq \mathbb{E} \left| \tilde{J}_\epsilon(0, u) \right| + 2 \|a\|_\infty \epsilon^{2-k}.$$

Now let $Q > 0$ and write $\Upsilon_\epsilon = \Upsilon_\epsilon^{Q,1} + \Upsilon_\epsilon^{Q,2}$ where

$$\Upsilon_\epsilon^{Q,1} = \Upsilon_\epsilon \mathbf{1}_{B_\epsilon(Q)}, \quad \Upsilon_\epsilon^{Q,2} = \Upsilon_\epsilon \mathbf{1}_{B_\epsilon(Q)^c}, \quad B_\epsilon(Q) = \left\{ \max_{[0,1]} |\tilde{x}_\epsilon| \leq Q \right\}.$$

Since $|\tilde{x}_\epsilon - W_\epsilon| \leq |\xi| + [t\epsilon^{-2}]\epsilon^2\|a\|_\infty + (t\epsilon^{-2} - [t\epsilon^{-2}])\epsilon\|v\|_\infty \leq |\xi| + \|a\|_\infty$ we have for $Q > 2(|\xi| + \|a\|_\infty + \|v\|_\infty)$ that

$$\begin{aligned} \mu[\max_{[0,1]} |\Upsilon_\epsilon^{Q,2}| > 0] &\leq \mu[\max_{[0,1]} |\tilde{x}_\epsilon| \geq Q] \leq \mu[\max_{[0,1]} |W_\epsilon| > Q/2] \\ &= \mathbb{P}[\max_{[0,1]} |W| > Q/4] + \left(\mu[\max_{[0,1]} |W_\epsilon| > Q/2] - \mathbb{P}[\max_{[0,1]} |W| > Q/4] \right). \end{aligned} \quad (9.12)$$

From Lemma 8.11 we have that

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq Q/4] \leq \frac{8\sqrt{2}\sigma}{Q\sqrt{\pi}} e^{-\frac{Q^2}{32\sigma^2}}.$$

The arguments between equations (8.23) and (8.25) in the proof of Theorem 8.2 apply verbatim. That is, it holds true that for $8\sqrt{2}\sigma/\sqrt{\pi} < Q_\epsilon = (-32\sigma^2 \log \epsilon)^{1/2}$ we have

$$\mathbb{P}[\max_{s \in [0,1]} |W(s)| \geq Q_\epsilon/4] \leq \epsilon.$$

By Assumptions 9.2 we have the General Framework **(H)** at hand and this implies Theorem 7.3 which gives that $\pi_1(W_\epsilon, W) \leq C\epsilon^{1/2-\delta}$ for all $\delta > 0$.

Hence we have that

$$\mu[\max_{[0,1]} |W_\epsilon| > Q_\epsilon/2] - \mathbb{P}[\max_{[0,1]} |W| > Q_\epsilon/4] \leq 2C\epsilon^{1/2-\delta}.$$

We deduce from equation (9.12) that

$$\mu[\max_{[0,1]} |\Upsilon_\epsilon^{Q_\epsilon, 2}| > 0] < \epsilon + 2C\epsilon^{1/2-\delta} = \mathcal{O}(\epsilon^{1/2-\delta}).$$

We now deal with the term $\Upsilon_\epsilon^{Q_\epsilon, 1} = \Upsilon_\epsilon \mathbb{1}_{B_\epsilon(Q_\epsilon)}$. Consider a grid S_ϵ on $[-Q_\epsilon, Q_\epsilon] \subset \mathbb{R}$ such that for all $x_\epsilon(n\epsilon^{k-2})$ there exists $u \in S_\epsilon$ with $|u - x_\epsilon(n\epsilon^{k-2})| \leq 2Q_\epsilon/|S_\epsilon|$. Then

$$|\tilde{a}(u, y(j)) - \tilde{a}(x_\epsilon(n\epsilon^k), y(j))| \leq \text{Lip}(\tilde{a}) \|u - x_\epsilon(n\epsilon^{k-2})\|_\infty \leq \frac{4LQ_\epsilon}{|S_\epsilon|}. \quad (9.13)$$

Since

$$|J_\epsilon(n)| \leq |\tilde{J}_\epsilon(n, u)| + |J_\epsilon(n) - \tilde{J}_\epsilon(n, u)|,$$

we get from (9.13) that for all $n \geq 0$ and $\epsilon > 0$

$$\mathbb{1}_{B_\epsilon(Q_\epsilon)} |J_\epsilon(n)| \leq \sum_{u \in S_\epsilon} |\tilde{J}_\epsilon(n, u)| + \frac{4LQ_\epsilon}{|S_\epsilon|}.$$

Hence by Minkowski's inequality we have

$$\begin{aligned} \left\| \max_{[0,1]} |\Upsilon_\epsilon^{Q_\epsilon, 1}| \right\|_p &\leq \sum_{n=0}^{[\epsilon^{-k}]-1} \epsilon^k \left\| \sum_{u \in S_\epsilon} |\tilde{J}_\epsilon(n, u)| \right\|_p + \frac{4LQ_\epsilon}{|S_\epsilon|} \\ &\leq \sum_{n=0}^{[\epsilon^{-k}]-1} \epsilon^k \left\| \sum_{u \in S_\epsilon} |\tilde{J}_\epsilon(0, u)| \right\|_p + 2\epsilon^{2-k} |S_\epsilon| \|a\|_\infty + \frac{4LQ_\epsilon}{|S_\epsilon|} \\ &\leq \sum_{u \in S_\epsilon} \left\| \tilde{J}_\epsilon(0, u) \right\|_p + 2\epsilon^{2-k} |S_\epsilon| \|a\|_\infty + \frac{4LQ_\epsilon}{|S_\epsilon|}. \end{aligned}$$

Notice that for each y fixed, we have that $A_u(y)$ is mean-zero and Lipschitz-continuous. Since $T : \Lambda \rightarrow \Lambda$ satisfies the General Framework **(H)** we can apply

Corollary 5.6 (with $A_u(y)$ playing the role of v) to obtain that $\left\| \tilde{\mathcal{J}}_\epsilon(0, u) \right\|_p \leq C_p \text{Lip}(\tilde{a}) \epsilon^{(2-k)/2}$. Hence there exists a constant $C' > 0$ such that

$$\left\| \max_{[0,1]} |\Upsilon_\epsilon^{Q_\epsilon, 1}| \right\|_p \leq C' \left(|S_\epsilon| \epsilon^{(2-k)/2} + \epsilon^{2-k} |S_\epsilon| + \frac{Q_\epsilon}{|S_\epsilon|} \right).$$

Choosing $|S_\epsilon| \approx \epsilon^{(k-2)/4}$ gives

$$\left\| \max_{[0,1]} |\Upsilon_\epsilon^{Q_\epsilon, 1}| \right\|_p \leq C \epsilon^{-\delta' + (2-k)/4}$$

where $\delta' > 0$ is arbitrarily small. To summarise we have

$$\begin{aligned} |Z_\epsilon| &= |Z_\epsilon(N\epsilon^k)| + |H_\epsilon| \leq |\Upsilon_\epsilon + I_\epsilon| + \mathcal{O}(\epsilon^k) \\ &\leq |\Upsilon_\epsilon| + \mathcal{O}(\epsilon^{k-1}) + \mathcal{O}(\epsilon^k) \leq |\Upsilon_\epsilon^{Q_\epsilon, 1}| + |\Upsilon_\epsilon^{Q_\epsilon, 2}| + \mathcal{O}(\epsilon^{k-1}). \end{aligned}$$

Since $k \in (1, 2)$ we obtain that

$$\begin{aligned} \left\| \max_{t \in [0,1]} |Z_\epsilon| \right\|_p &\leq \left\| \max_{t \in [0,1]} |\Upsilon_\epsilon^{Q_\epsilon, 1}| \right\|_p + \left\| \max_{t \in [0,1]} |\Upsilon_\epsilon^{Q_\epsilon, 2}| \right\|_p + \mathcal{O}(\epsilon^{k-1}) \\ &\leq C \epsilon^{-\delta' + (2-k)/4} + \mathcal{O}(\epsilon^{1/2-\delta}) + \mathcal{O}(\epsilon^{k-1}). \end{aligned}$$

This is optimal at $k = 6/5$ and we obtain the desired result. ■

Proof of Theorem 9.1: Recall from Lemma 9.3 that $\left| \hat{Z}_\epsilon(t) - Z_\epsilon(t) \right| \leq C \epsilon^{1/5}$. By Proposition 9.4 we obtain that

$$\begin{aligned} \left\| \max_{t \in [0,1]} \left| (W_\epsilon + \hat{Z}_\epsilon) - W_\epsilon \right| \right\|_p &\leq \left\| \max_{t \in [0,1]} |Z_\epsilon| \right\|_p + \mathcal{O}(\epsilon^{1/5}) \\ &= \mathcal{O}(\epsilon^{(\frac{1}{5}-\delta')}). \end{aligned}$$

It follows from Proposition 7.4 that there exists a real constant $C > 0$ such that

$$\pi_1(W_\epsilon + \hat{Z}_\epsilon, W_\epsilon) \leq C \epsilon^{(\frac{1}{5}-\delta') \frac{p}{p+1}}.$$

Define the continuous map $\mathcal{G} : \mathcal{C}([0, 1], \mathbb{R}) \rightarrow \mathcal{C}([0, 1], \mathbb{R})$ given by $\mathcal{G}(u) = v$ where $v(t) = \xi + u(t) + \int_0^t \bar{a}(v(s)) ds$. By Theorem 8.8, \mathcal{G} is a well-defined, Lipschitz continuous map and

$$\|\mathcal{G}(u_1) - \mathcal{G}(u_2)\|_\infty \leq \text{Lip}(\mathcal{G}) \|u_1 - u_2\|_\infty .$$

Thus we can apply the Lipschitz Mapping Theorem (Theorem 7.9) to obtain

$$\pi_1(\hat{x}_\epsilon, X) = \pi_1(\mathcal{G}(W_\epsilon + \hat{Z}_\epsilon), \mathcal{G}(W)) \leq (\text{Lip}(\mathcal{G}) \vee 1) \pi_1(W_\epsilon + \hat{Z}_\epsilon, W) .$$

Since

$$\pi_1(W_\epsilon + \hat{Z}_\epsilon, W) \leq \pi_1(W_\epsilon, W) + \pi_1(W_\epsilon + \hat{Z}_\epsilon, W_\epsilon)$$

we conclude that $\pi_1(\hat{x}_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$. ■

9.3 Rates of Convergence in the Multiplicative Noise case for Maps

We now turn to investigating the rate of convergence for maps in the presence of multiplicative noise.

Theorem 9.5. (Main Theorem for Maps) *Consider equation (9.2) and let X denote solutions of the SDE of the form*

$$dX = b(X) \circ dW + (\bar{a}(X) - \frac{1}{2}b(X)b'(X) \int_\Lambda v^2 d\mu) dt, \quad X(0) = \xi. \quad (9.14)$$

Let Assumptions 9.2 hold. Also suppose that $b \in \mathcal{C}(\mathbb{R})$, b , b' , b'' , and $1/b$ are bounded and that b and b' are uniformly Lipschitz continuous. Then $\pi_1(\hat{x}_\epsilon, X) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$.

Proof: Define $\theta_\epsilon(n) = \psi(x_\epsilon(n))$ and set $\psi' = 1/b$. By Taylor's Theorem we have that

$$\begin{aligned}\theta_\epsilon(n+1) - \theta_\epsilon(n) &= \psi(x_\epsilon(n+1)) - \psi(x_\epsilon(n)) \\ &= \psi'(x_\epsilon(n))(x_\epsilon(n+1) - x_\epsilon(n)) + \frac{1}{2}\psi''(x_\epsilon(n))(x_\epsilon(n+1) - x_\epsilon(n))^2 \\ &\quad + \mathcal{O}(|x_\epsilon(n+1) - x_\epsilon(n)|^3)\end{aligned}\tag{9.15}$$

Substituting for $x_\epsilon(n)$ using equation (9.2) into equation (9.15) we obtain

$$\begin{aligned}\theta_\epsilon(n+1) - \theta_\epsilon(n) &= \psi'(x_\epsilon(n))[\epsilon b(x_\epsilon(n))v(y(n)) + \epsilon^2 a(x_\epsilon(n), y(n), 0) \\ &\quad + \epsilon^2(a(x_\epsilon(n), y(n), \epsilon) - a(x_\epsilon(n), y(n), 0))] \\ &\quad + \frac{1}{2}\psi''(x_\epsilon(n))\epsilon^2 b^2(x_\epsilon(n))v^2(y(n)) + \mathcal{O}(\epsilon^3).\end{aligned}$$

Using $\psi' = 1/b$ and by Assumptions 9.2 we have

$$\begin{aligned}\theta_\epsilon(n+1) - \theta_\epsilon(n) &= \epsilon v(y(n)) + \epsilon^2[\psi'(x_\epsilon(n))a(x_\epsilon(n), y(n), 0) \\ &\quad + \frac{1}{2}\psi''(x_\epsilon(n))[(\psi'(x_\epsilon(n)))^{-2}v^2(y(n))]] + \mathcal{O}(\epsilon^{11/5}) \\ &= \epsilon v(y(n)) + \epsilon^2 \tilde{f}(\theta_\epsilon(n), y(n), \epsilon),\end{aligned}$$

where

$$\begin{aligned}\tilde{f}(\theta, y, \epsilon) &= \psi'(\psi^{-1}(\theta))a(\psi^{-1}(\theta), y, 0) \\ &\quad + \frac{1}{2}\psi''(\psi^{-1}(\theta))[\psi'(\psi^{-1}(\theta))]^{-2}v(y)^2 + \mathcal{O}(\epsilon^{1/5}).\end{aligned}$$

We claim that \tilde{f} is bounded and uniformly Hölder. By Assumptions 9.2 we have that a is bounded and uniformly Lipschitz and v is bounded and uniformly Lipschitz. Moreover, by assumptions of the Theorem we have that ψ' is uniformly Lipschitz since it is bounded and b is uniformly Lipschitz. Since b is also bounded we have by the Inverse Function Theorem that $(\psi^{-1})'$ is also bounded and by the Mean Value Theorem we obtain that ψ^{-1} is uniformly Lipschitz. Lastly, notice that $\psi'' = -b'/b^2$ where by assumption b' is bounded

and uniformly Lipschitz. Thus we conclude that \tilde{f} is bounded and uniformly Hölder, being the product and composition of bounded, uniformly Hölder continuous functions.

Let

$$\tilde{F}(\theta) = \psi'(\psi^{-1}(\theta))\bar{a}(\psi^{-1}(\theta)) + \frac{1}{2}\psi''(\psi^{-1}(\theta))[\psi'(\psi^{-1}(\theta))]^{-2} \int_{\Lambda} v^2 d\mu$$

and observe that $\tilde{F}(\theta) = \int_{\Lambda} \tilde{f}(\theta, y, 0) d\mu(y)$. Consider the SDE

$$d\Theta = dW + \tilde{F}(\Theta)dt, \quad \Theta(0) = \psi(\xi). \quad (9.16)$$

Then it follows from Theorem 9.1 that the rate of convergence of θ_ϵ where

$$\theta_\epsilon(n+1) = \theta_\epsilon(n) + \epsilon v(y(n)) + \epsilon^2 \tilde{f}(\theta_\epsilon, y(n), \epsilon), \quad \theta(0) = \psi(\xi),$$

to Θ , where Θ is a solution of the SDE (9.16), is $\pi_1(\theta_\epsilon, \Theta) = \mathcal{O}(\epsilon^{1/5-\delta})$ for all $\delta > 0$. Next, it is immediate that

$$\begin{aligned} \tilde{F}(\Theta) &= \frac{\bar{a}(X)}{b(X)} + \frac{-b'(X)}{2b^2(X)}(b^2(X) \int_{\Lambda} v^2 d\mu) \\ &= \frac{\bar{a}(X)}{b(X)} - \frac{1}{2}b'(X) \int_{\Lambda} v^2 d\mu. \end{aligned}$$

Recall that the chain rule in Stratonovich's Theory satisfies the usual laws of ordinary deterministic calculus and note that

$$(\psi^{-1})'(\Theta) = (\psi^{-1})'(\psi(X)) = \frac{1}{\psi'(X)} = b(X)$$

by a straightforward application of the Inverse Function Theorem. Hence using the SDE (9.14) we have that the limiting process $X = \psi^{-1}(\Theta)$ is given by

$$\begin{aligned} dX &= (\psi^{-1}(\Theta))' \circ d\Theta = b(X) \circ (dW + \frac{\bar{a}(X)}{b(X)} - \frac{1}{2}b'(X) \int_{\Lambda} v^2 d\mu)dt \\ &= b(X) \circ dW + [\bar{a}(X) - \frac{1}{2}b(X)b'(X) \int_{\Lambda} v^2 d\mu]dt. \end{aligned}$$

Also, $X(0) = \psi^{-1}(\Theta(0)) = \psi^{-1}(\psi(\xi)) = \xi$. Thus we have, as at the end of the proof of Theorem 8.12, that we can apply the Lipschitz Mapping Theorem to conclude that

$$\pi_1(\hat{x}_\epsilon, X) = \pi_1(\psi^{-1}(\theta_\epsilon), \psi^{-1}(\Theta)) \leq (\text{Lip}(\psi^{-1}) \vee 1)\pi_1(\theta_\epsilon, \Theta) = \mathcal{O}(\epsilon^{1/5-\delta})$$

for all $\delta > 0$. ■

Bibliography

- [1] Antoniou M. & Melbourne I., *Rate of convergence in the Weak Invariance Principle for deterministic systems*. In preparation.
- [2] Arak T. V., *On an estimate due to A. A. Borovkov*. Teor. Verojatnost. i Primenen., 20: 380–381, 1975.
- [3] Baladi V., *Positive transfer operators and decay of correlations*. Advanced Series in Nonlinear Dynamics, vol. 16, World Scientific Publishing Co. Inc., River Edge, NJ, 2000.
- [4] Bass R. F., *Stochastic processes*. Cambridge Series in Statistical and Probabilistic Mathematics, 33. Cambridge University Press, Cambridge, 2011.
- [5] Berry A. C., *The accuracy of the Gaussian approximation to the sum of independent variates*. Trans. Amer. Math. Soc., 49: 122-136, 1941.
- [6] Billingsley P., *The invariance principle for dependent random variables*. Trans. Amer. Math. Soc., 83: 250-268, 1956.
- [7] Billingsley P., *The Lindeberg-Lévy theorem for martingales*. Proc. Amer. Math. Soc., 12: 788-792, 1961.
- [8] Billingsley P., *Probability and measure*. 3rd edition, Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995.

- [9] Billingsley P., *Convergence of probability measures*. 2nd edition, Wiley Series in Probability and Statistics: Probability and Statistics, A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1999.
- [10] Birkhoff G. D., *Proof of the ergodic theorem*. Proc. Natl. Acad. Sci., 17: 656-660, 1931.
- [11] Bolthausen E., *Exact convergence rates in some martingale Central Limit Theorems*. Ann. Probab., 10(3): 672-688, 1982.
- [12] Borovkov A. A., *On the rate of convergence in the invariance principle*. Probab. Theory Appl., 18: 217-234, 1973.
- [13] Boyarsky A. & Góra P., *Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension*. Probability and its Applications, Birkhäuser Boston Inc., Boston, 1997.
- [14] Breiman L., *Probability*. Corrected reprint of the 1968 original. Classics in Applied Mathematics, 7. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1992.
- [15] Brown B. M., *Martingale Central Limit Theorems*. Ann. Math. Statist., 42: 59-66, 1971.
- [16] Chow Y. S. & Teicher H. *Probability Theory. Independence, Interchangeability, Martingales*. Reprinted 3rd edition, Springer Texts in Statistics, Springer-Verlag, New York, 1997.
- [17] Burkholder D. L., *Distribution function inequalities for martingales*. Ann. Probability, 1: 19-42, 1973.
- [18] Dolgopyat D., *Limit theorems for partially hyperbolic sets*. Trans. Amer. Math. Soc., 356: 1637-1689, 2004.
- [19] Dolgopyat D., *Averaging and invariant measures*. Mosc. Math. J., 5: 537-576, 2005.

- [20] Donsker M. D., *An invariance principle for certain probability limit theorems*. Mem. Amer. Math. Soc., 6, 1951.
- [21] Dudley R. M., *Convergence of Baire measures*. Studia Math., 27: 251–268, 1966.
- [22] Dudley R. M., *Distances of probability measures and random variables*. Ann. Math. Statist 39: 1563–1572, 1968.
- [23] Dudley R. M., *Probabilities and metrics. Convergence of laws on metric spaces, with a view to statistical testing*. Lecture Notes Series, No. 45. Matematisk Institut, Aarhus Universitet, Aarhus, 1976.
- [24] Dudley R. M., *Real Analysis and Probability*. Revised reprint of the 1989 original, Cambridge Studies in Advanced Mathematics, 74, Cambridge University Press, Cambridge, 2002.
- [25] Durrett R., *Probability: Theory and Examples*. 4th edition, Cambridge Series in Statistical and Probabilistic Mathematics, 31, Cambridge University Press, Cambridge, 2010.
- [26] Dvoretzky A., *Asymptotic normality for sums of dependent random variables*. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 513–535. Univ. California Press, Berkeley, Calif., 1972.
- [27] Esseen C.-G., *On the Liapunoff limit of error in the theory of probability*. Ark. Mat. Astron. Fys., 28A: 1-19, 1942.
- [28] Ethier S. N. & Kurtz T. G., *Markov processes. Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.
- [29] Field M., Melbourne I. & Török A., *Decay of Correlations, Central Limit Theorems and approximation by Brownian Motion for Compact Groups Actions*. Ergod. Th. & Dynam. Sys., 23: 87-110, 2003.

- [30] Gibbs A. L. & Su F. E., *On choosing and bounding probability metrics*. Internat. Statist. Rev., 70: 419-435, 2002.
- [31] Gordin M. I., *The Central Limit Theorem for stationary processes*. Soviet Math. Dokl., 10: 1174-1176, 1969.
- [32] Gottwald G. A. & Melbourne I., *Homogenization for deterministic maps and multiplicative noise*. Proc. R. London A, 20130201, 2013.
- [33] Gouëzel S., *Berry–Esseen theorem and local limit theorem for non uniformly expanding maps*. Ann. Inst. H. Poincaré Probab. Statist., 41: 997-1024, 2005.
- [34] Gouëzel S., *Limit theorems in dynamical systems using the spectral method*. Proc. Sympos. Pure Math., 89, Amer. Math. Soc., Providence, RI, 2015.
- [35] Grama I., Le Page É. & Peigné M., *On the rate of convergence in the Weak Invariance Principle for dependent random variables with applications to Markov chains*. Colloq. Math., 134: 1–55, 2014.
- [36] Hall P. & Heyde C. C., *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics, Academic Press, Inc., New York-London, 1980.
- [37] Haeusler E., *On the rate of convergence in the Central Limit Theorem for martingales with discrete and continuous time*. Ann. Probab., 16(1): 275–299, 1988.
- [38] Hofbauer F. & Keller G., *Ergodic properties of invariant measures for piecewise monotonic transformations*. Math. Z., 180: 119–140, 1982.
- [39] Huber P. J., *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1981.
- [40] Ibragimov I. A., *A Central Limit Theorem for a class of dependent random variables*. Theory Probab. Appl., 8: 83-89, 1963.

- [41] Ikeda N. & Watanabe S., *Stochastic differential equations and diffusion processes*. 2nd edition. North-Holland Mathematical Library, 24. North-Holland Publishing Co., Amsterdam, Kodansha, Ltd., Tokyo, 1989.
- [42] Jacod J. & Shiryaev A. N., *Limit theorems for stochastic processes*. 2nd edition, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 288, Springer-Verlag, Berlin, 2003.
- [43] Joos K., *Abschätzungen der Konvergenzgeschwindigkeit in asymptotischen Verteilungsaussagen für Martingale*. Ph.D. thesis, LMU, Ludwig-Maximilians-Universität, München, 1989.
- [44] Kallenberg O., *Foundations of modern probability*. 2nd edition, Probability and its Applications (New York), Springer-Verlag, New York, 2002.
- [45] Karatzas I. & Shreve S. E., *Brownian motion and stochastic calculus*. 2nd edition. Graduate Texts in Mathematics, 113. Springer-Verlag, New York, 1991.
- [46] Katok A. & Hasselblatt B., *Introduction to the modern theory of dynamical systems*. With a supplementary chapter by Katok and Leonardo Mendoza. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, 1995.
- [47] Kelly D. & Melbourne I., *Smooth approximation of stochastic differential equations*. Ann. Probab., 44: 479–520, 2016.
- [48] Kelly D. & Melbourne I., *Deterministic homogenization for fast-slow systems with chaotic noise*. J. Funct. Anal., 272: 4063–4102, 2017.
- [49] Kolmogorov A., *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York, 1956. [Translation from german of *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933.]
- [50] Koopman B. O., *Hamiltonian Systems and Transformations in Hilbert Space*. Proc. Nat. Acad. Sci., 17: 315–318, 1931.

- [51] Korepanov A., Kosloff Z. & Melbourne I., *Martingale-coboundary decomposition for families of dynamical systems*. Preprint August 2016. To appear in Ann. Inst. H. Poincaré Anal. Non Linéaire.
- [52] Kubilius K., *Rate of convergence in the invariance principle for martingale difference arrays*. Lithuanian Math. J., 34: 383-392, 1994.
- [53] Machkouri El M. & Ouchti L., *Exact Convergence rates in the Central Limit Theorem for a Class of Martingales*. Bernoulli, 13: 981-999, 2007.
- [54] Mann H. B. & Wald A., *On Stochastic Limit and Order Relationships*. Ann. Math. Statistics, 14: 217–226, 1943.
- [55] Melbourne I. & Stuart A., *A note on diffusion limits of chaotic skew product flows*. Nonlinearity, 24: 1361-1367, 2011.
- [56] McLeish D. L., *Dependent Central Limit Theorems and Invariance Principles*. Ann. Probab., 2: 620-628, 1974.
- [57] McLeish D. L., *Invariance principles for dependent variables*. Springer-Verlag, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 32: 165-178, 1975.
- [58] Mourrat J.-C., *On the rate of convergence in the martingale Central Limit Theorem*. Bernoulli, 19: 633–645, 2013.
- [59] Neumann J. V., *Proof of the quasi ergodic hypothesis*. Proc. Nat. Acad. Sci., 18: 70-82, 1932.
- [60] Parry W. & Pollicott M., *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics*. Astérisque, 187-188, Société Mathématique de France, Montrouge, 1990.
- [61] Parthasarathy K. R., *Probability measures on metric spaces*. Probability and Mathematical Statistics, 3, Academic Press, Inc., New York-London, 1967.

- [62] Pavliotis G. A., *Stochastic processes and applications. Diffusion processes, the Fokker-Planck and Langevin equations*. Texts in Applied Mathematics, 60. Springer, New York, 2014.
- [63] Pavliotis G. A. & Stuart A. M., *Multiscale Methods: Homogenization and Averaging*. Texts in Applied Mathematics, 53, Springer, New York, 2008.
- [64] Petersen K., *Ergodic theory*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1989.
- [65] Pollard D., *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.
- [66] Pollicott M. & Yuri M., *Dynamical Systems and Ergodic Theory*. London Mathematical Society Student Texts, 40, Cambridge University Press, Cambridge, 1998.
- [67] Prokhorov Yu. V., *Convergence of Random Processes and Limit Theorems in Probability Theory*. Theory Probab. Appl., 1: 157–214, 1956.
- [68] Ratner M., *The central limit theorem for geodesic flows on n -dimensional manifolds of negative curvature*. Israel J. Math., 16: 181–197, 1973.
- [69] Revuz D. & Yor M., *Continuous martingales and Brownian motion*. 3rd edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 293, Springer-Verlag, Berlin, 1999.
- [70] Rogers L. C. G. & Williams D., *Diffusions, Markov processes, and martingales. Vol. 2. Itô calculus*. Reprint of the 2nd (1994) edition, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000.
- [71] Rohlin V. A., *Exact endomorphisms of a Lebesgue space*. Izv. Akad. Nauk SSSR Ser. Mat., 25: 499–530, 1961.
- [72] Rohlin V. A., *Selected topics from the metric theory of dynamical systems*. Uspekhi Mat. Nauk, 4: 57–128, 1949. [A.M.S. Transl. Ser. 2, 49, 1966.]

- [73] Royden H. & Fitzpatrick P., *Real Analysis*, Fourth edition, Pearson Education Inc., Boston, 2010.
- [74] Schilling R. L., *Measures, integrals and martingales*. 2nd edition. Cambridge University Press, Cambridge, 2017.
- [75] Schilling R. L. & Partzsch L., *Brownian motion. An introduction to stochastic processes*. With a chapter on simulation by Björn Böttcher. 2nd edition. De Gruyter Graduate. De Gruyter, Berlin, 2014.
- [76] Sinai J. G., *The Central Limit Theorem for geodesic flows on manifolds of constant negative curvature*. Soviet Math. Dokl., 1: 983–987, 1960.
- [77] Skorokhod A. V., *Limit Theorems for Stochastic Processes*. Theory Probab. Appl., 1(3): 261–290, 1956.
- [78] Stone C., *Weak convergence of stochastic processes defined on semi-infinite time intervals*. Proc. Amer. Math. Soc., 14: 694–696, 1963.
- [79] Strassen V., *The existence of probability measures with given marginals*. Ann. Math. Statist., 36: 423–439, 1965.
- [80] Stratonovich R. L., *A new representation for stochastic integrals and equations*. SIAM J. Control 4: 362–371, 1966.
- [81] Sussmann H. J., *On the gap between deterministic and stochastic ordinary differential equations*. Ann. Probability 6: 19–41, 1978.
- [82] Sussmann H. J., *Limits of the Wong-Zakai type with a modified drift term*. Stochastic analysis, 475–493, Academic Press, Boston, MA, 1991.
- [83] van der Vaart A. W. & Wellner J. A., *Weak convergence and empirical processes, with applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York, 1996.
- [84] Varadarajan V. S., *Weak convergence of measures on separable metric spaces*. Sankhyā, 19: 15–22, 1958.

- [85] Volný D., *Approximating martingales and the Central Limit Theorem for strictly stationary processes*. Stochastic Process. Appl., 44: 41-74, 1993.
- [86] Whitt W., *Weak convergence of probability measures on the function space $C[0, \infty)$* . Ann. Math. Statist., 41: 939-944, 1970.
- [87] Whitt W., *Preservation of rates of convergence under mappings*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 29: 39-44, 1974.
- [88] Whitt W., *Stochastic-Process Limits: An introduction to stochastic-process limits and their application to queues*. Springer Series in Operations Research, Springer-Verlag, New York, 2002.
- [89] Williams D., *Probability with martingales*. Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.
- [90] Wong E. & Zakai M., *On the convergence of ordinary integrals to stochastic integrals*. Ann. Math. Statist., 36: 1560-1564, 1965.
- [91] Wong E. & Zakai M., *Riemann-Stieltjes approximations of stochastic integrals*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 12: 87-97, 1969.
- [92] Zolotarev V. M., *Probability metrics*. Theory Probab. Appl., 28: 278-302, 1984.

